

Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225
 Category: similarity to known protein

1 MNQPCNSMEP RVMDDMLKL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL
 51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGLNLAHLV WLDLSFNIE
 101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
 151 YLRRFKCLRT LSLSRNPISE AEDYKMFICA YLPDLMYLDY RRIDDHTASV
 201 SLSVSQPCET DSSSPQVSWK RGIEE

BLASTP hits

Entry S68209 from database PIR:
 sds22 protein homolog - human >TREMBL:HSSDS22MR_1 gene: "sds22";
 product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA
 Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:
 suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)
 >SWISSPROT:SD22 SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT
 SDS22. >TREMBL:SPAC4A8_12 gene: "sds22"; product: "phosphatases pp1
 regulatory subunit"; S.pombe chromosome I cosmid c4A8.
 Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:
 protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)
 >SWISSPROT:SD22 SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT
 SDS22. >TREMBL:SPAC4A8_12 gene: "sds22"; product: "phosphatases pp1
 regulatory subunit"; S.pombe chromosome I cosmid c4A8.
 Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2_5 from database TREMBL:
 gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.
 Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

Alert BLASTP hits for DKFZphut1_20ml1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_20ml1, frame 1

Report for DKFZphut1_20ml1.1

[LENGTH] 225
 [MW] 25955.87
 [pI] 4.63
 [HOMOL] PIR:S68209 sds22 protein homolog - human 1e-18
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
 [FUNCAT] 01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
 YJL005w] 3e-05
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
 [FUNCAT] 10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
 [EC] 4.6.1.1 Adenylate cyclase 2e-06
 [PIRKW] nucleus 5e-16
 [PIRKW] duplication 2e-06
 [PIRKW] tandem repeat 2e-06
 [PIRKW] cAMP biosynthesis 2e-06
 [PIRKW] glycoprotein 2e-06
 [PIRKW] phosphorus-oxygen lyase 2e-06
 [SUPFAM] leucine-rich alpha-2-glycoprotein repeat homology 5e-16
 [SUPFAM] fibromodulin 3e-07
 [SUPFAM] yeast adenylate cyclase catalytic domain homology 2e-06
 [SUPFAM] yeast adenylate cyclase 2e-06
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 1

[KW] All_Alpha

```

SEQ  MNQPCNSMEPRVMDDDLKLA VGDQGPQEEAGQLAKQEGILFKDVLSLQLDFRNILRIDN
PRD  cccccccccccccchhhhhccccccccchhhhhhhhhhhchhhhhhhhhccccccccccccc

SEQ  LWQFENLRKLQLDNNIEKIEGLENL AHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLQVLSLGNNRIDNMMNIIYLRREFKCLRTLSLSRNPISEAEDYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccccccchhhhhhhccccccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVLSVSPQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccccc

```

Prosite for DKFZphut1_20m11.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1_20m11.1)

DKFZphute1_20m24

group: metabolism

DKFZphute1_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits
Alg9 is involved in the assembly of the core oligosaccharide
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)

HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```
1 TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51 TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCGGC TCGGACAAG
101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GGCGGCGCG AGCACCGGAC
151 CGAGTTATCT GGGAAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGCTGCTT TCAGCAAGGT TATGTGCTGC TCTCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTCGGTC CTATGCTTAC CTGTTGCTTC ATGCCTGGCC AGCTGCATT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTTGTGTTT ACTTTTGCG
451 ATGTCTTCTG GCTTTTGTA GCTGTATTG TGAACCTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATGTCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTTG ATTTGCTGGT
751 CATGAAACAC AGGTGGAAGA GTTTCTTTCA TTGGTCGCTG ATGGCCCTCA
801 TACTATTCTT GGTGCCTGTG GTGGTCATG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTAATCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCTCG GTATTCTAT TTAATTAATG
951 GATTCTTGAA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAATTT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTCTTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCCTTT CCCTGTGTAT
1151 CCCATTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTTGTCATT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTTGTA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGA TCGATTTCCT
1451 AGCAGCTTCC TTCTTCCTGA CAATTGGCAG CTTCAGTTCA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTCATC CAATAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTC CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCCTT CCTGTCAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAACCCC GGAAGGCAAA GCAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATCGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSAC381 from database EMBL:
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.
Length = 601

Medline entries

96293493:
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: identification of the ALG9 gene encoding a putative mannosyl transferase.

Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611
Category: strong similarity to known protein

```

1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTLSGNKA
51 GQVWAPGEGT AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEFG
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVEY FLRCLLAFVS
151 CICELYFYKA VCKKFGHLVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPFs AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVFL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVOLD TMRETPREFK
551 YSSNKEEWS LAYRPFLDAS RSSKLLRAFY VPFLSDQYTV YVNYTILKPR
601 KAKQIRKKSG G

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20m24, frame 2

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.

Length = 653

HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query:  48 NKAGQVWAPGEGTAFKCLLSARLCAALLSNISDCDETFNYWEPH YLIYGEFGQTWEYSP 107
      N   W   + FK LLS R+  A+  I+DCDE +NYWEP H  +YGEFGQTWEYSP
Sbjct:  43 NNPDNDWPFSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFGQTWEYSP 102

Query:  108 AYAIRSYAYLLHAWPAAFHARILQTNKILVEYFLRCLLAFVSCICELYFYKAVCKKFG 167
      YAIRSY Y+ LH  PA+  A+  KI+VF +R  +  + E Y + A+CKK  +
Sbjct:  103 VYAIRSYFYIYLHYIPASLFANLFGDTRIVVFTLIRLTIGLFCLLGEYAFDAICKKINI 162

Query:  168 HVSRMMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
      R  + F + S+GMF +S+AF+PSSFCM T  +  +  +  +  + VA  ++GW
Sbjct:  163 ATGRFFILFSIFSSGMFLASTAFVPSSFCMAITFYILGAYLNENWTAGIFCVAFSTMVGW 222

Query:  228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```

Sbjct: 223 PFSAVLGLPIVADMLLLKGLRIRFILTSLVIGLCIGGVQVITDSHYFGKTVLAPLNIFLY 282

Query: 288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFALALLVPLTSLMEYLLQRFHVQNLGHPY 347
 NV + GP LYG EP FY+ N F N+N+ A PL+ + Y + + Q+

Sbjct: 283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340

Query: 348 WLTAPMYI-----WFIIFFIQPHKEERFLFPVYPLICGAVALSALQKCYHFVFOR 400
 + AP+ + W +IF Q HKEERFLFP+YP I A+AL A + ++

Sbjct: 341 YQRFAPILLAVTTAAWLLIFGSAHKEERFLFPYFFIAFFAALALDATNR---LCLKK 397

Query: 401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPE 460
 ++ N L++ + F +LS SR+ ++ Y +++Y T+ T +

Sbjct: 398 LGMD-----NILSILFILCFAILSASRTYSIHNNYGSVHEIYRSLNAELTNRT-NFKNF 450

Query: 461 GRPVNVCGKWEYRFPSSFLPDNW-----QLQFIPSEFRGQLPKPFAEGPL---ATRI 511
 P+ VCVGKEW+RFPSSF +P +++FI SEFRG LKPKF + TR

Sbjct: 451 HDPIRVCVGKEWHRFPSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPFLLKSDKLVEVTRH 510

Query: 512 VPTDMNDQNLEEPSRYIDISKCHYLVDLDTMRETREPKEYSSNKEEW 558
 +PT+MN+ N EE SRY+D+ C Y+VD+D M ++ REP + ++ +

Sbjct: 511 IPTMNNLNQEEISRYVDLSDCDYVVDVD-MPQSDREPDFRKMQRNY 556

Pedant information for DKFZphutel_20m24, frame 2

Report for DKFZphutel_20m24.2

[LENGTH] 611
 [MW] 69863.78
 [pI] 8.91
 [HOMOL] SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-93
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69
 [PIRKW] glycosyltransferase 9e-68
 [PIRKW] transmembrane protein 9e-68
 [PIRKW] hexosyltransferase 9e-68
 [PROSITE] MYRISTYL 9
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] PKC_PHOSPHO_SITE 6
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] TRANSMEMBRANE 7
 [KW] LOW_COMPLEXITY 6.71 %

SEQ MASRGARQRLKSGASSGDTAPAADKLELLGSREAGGAHRTLSGNKAGQVWAPEGST
 SEG
 PRD ccchhhhhhhccccccccccccchhhhhhhhhccccccccccccccccccccccccch
 MEMMMMMMM

SEQ AFKCLLSARLCAALLSNISDCDETFFNYWEPHYLIYGEFGTWEYSPAYAIRSYAYLLLH
 SEGxxxxxxxxxxxxxxxxx.....
 PRD hhhhhhhhhhhhhhhhhhhhhccch
 MEM MM.....M

SEQ AWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAFLVLS
 SEG
 PRD cchhhcc
 MEM MM.....

SEQ TGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cccccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccch
 MEMMMMMMMMMMMMMMMMM

SEQ LLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHGPDLYGT
 SEG
 PRD hhhcc
 MEM MM.....

SEQ EPWYFYLINGFLNFNVAFALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cccccccccccccchhh
 MEMMM

```

SEQ      FFIQPHKEERFLFPVYPPLICLGAVALSALQKCYHFVFQRYRLEHYTVTSNWLALGTVFL
SEG
PRD      hhccccchhhhhhhccceeehhhhhhhhhhhhhhhhhhhhhhhhhhheeeccchhhhhhhhee
MEM      .....MMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMM.....

SEQ      FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVGKEWYRFPSSFL
SEG
PRD      eehhhhhhhheeeccccccccccccceeeccccccccceccccceeeeeecccccccccccc
MEM      .....

SEQ      LPDNWLQLFIPSEFRQLPKPFAEGPLATRIVPTDMNDQNLEEPSRYIDISKCHYLVLDL
SEG
PRD      cccccceeeccccccccccccccccceeeccccccccccccccccceeeeeeceeeeeeccc
MEM      .....

SEQ      TMRETPREPKYSSNKEEWISLAYRPFLDASRSKLLRAFYPVFLSDQYTVVYNYTILKPR
SEG
PRD      cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhheeeeeeeceeeeeeeeeeecccc
MEM      .....

SEQ      KAKQIRKKSGG
SEG
PRD      hhhhhhhcccc
MEM

```

Prosite for DKFZphute1_20m24.2

PS000001	77->81	ASN_GLYCOSYLATION	PDOC000001
PS000001	593->597	ASN_GLYCOSYLATION	PDOC000001
PS000004	606->610	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	67->70	PKC_PHOSPHO_SITE	PDOC000005
PS000005	133->136	PKC_PHOSPHO_SITE	PDOC000005
PS000005	541->544	PKC_PHOSPHO_SITE	PDOC000005
PS000005	545->548	PKC_PHOSPHO_SITE	PDOC000005
PS000005	553->556	PKC_PHOSPHO_SITE	PDOC000005
PS000005	572->575	PKC_PHOSPHO_SITE	PDOC000005
PS000006	16->20	CK2_PHOSPHO_SITE	PDOC000006
PS000006	79->83	CK2_PHOSPHO_SITE	PDOC000006
PS000006	329->333	CK2_PHOSPHO_SITE	PDOC000006
PS000006	457->461	CK2_PHOSPHO_SITE	PDOC000006
PS000006	541->545	CK2_PHOSPHO_SITE	PDOC000006
PS000006	545->549	CK2_PHOSPHO_SITE	PDOC000006
PS000006	553->557	CK2_PHOSPHO_SITE	PDOC000006
PS000008	12->18	MYRISTYL	PDOC000008
PS000008	14->20	MYRISTYL	PDOC000008
PS000008	32->38	MYRISTYL	PDOC000008
PS000008	47->53	MYRISTYL	PDOC000008
PS000008	166->172	MYRISTYL	PDOC000008
PS000008	182->188	MYRISTYL	PDOC000008
PS000008	218->224	MYRISTYL	PDOC000008
PS000008	222->228	MYRISTYL	PDOC000008
PS000008	234->240	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphute1_20m24.2)

DKFZphutel_21d15

group: uterus derived

DKFZphutel_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGCGCGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCTAGTG ACCGCGGCGC TCGCGGGCCT GCGGCGCGTT
251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC CGCGGCCCTT CCCCTGGGCG
301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GCGCGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCTCGA GGCCGCAGTG GCGCGCGCCA
401 GACCACTGCC AGGCTCAGGC GCGGCGCGGG CTGGGCGACG GCGAGGACGC
451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTC GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCGG
601 GCCCCAACAC CGTGCCCAAG GCCCGGGGCG CGAGCCCACC TTAGGTCCCC
651 TCACCCGGGT GGAGGGCATC AAGGTGAGGA CCTCCTTGCC CCGCCGCGCT
701 CCAGGCCCTG CACGGCTGAG CCCGAGAGGA CCGGCGCTCA GCGCGGGTCC
751 CCACGCTGCC CCCGGCGCTG CTCTGCGTCG GTCCCGCGCG CTCCCACTCA
801 CTCGCTGTCT GTCGCTCTCC GGGCCGGGGC GACTTGGCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCCGGCTT CCTGACTGAT
1001 GAAGAGTGTC GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTACAGCG
1051 CAGCCAGATC CTGCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTACAGCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGT TCTGGCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCGTG
1301 AGCCTCCTGC CCACTCCAG GTGCACAATT TTGAAAACCT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTG GTGATGGTCT CGAGGGCAGT TCTTGAGAC
1501 CCTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCTT GCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCCT TCCTCTGTG CACCTGGAGT CATGCCGAAG CGCCTAAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTCATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGA AGGTGATGTC CTTGGGTGAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTGTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTGTC TAAAGTGGCT GGTAAACGAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCC AGGATGAAC CTGGGGTCAT GAGGAACCTC
2301 CCGGGGGCTG GCCCTGCTTG CACCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGCCCATATG
2451 GCTCAAGCTT CTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAAGTCTAG AAAAGGAAC
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAAGTGA CAAGGCCCTT CAGTGCCTGA
2651 GGTCAAGGGC TGCTGCTGCC TTTGGGGTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCCT CCAATATAGG GCTTGGTGGG CATTGATGGT GGGTGCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTCCGGGCT
2801 TCCACAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGGCG
```

```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGCCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGCTCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCGGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCCTCTGT CCCACCAGGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCCG CTGCAGGTTG TTCGATATGG TGAGGGGGGC
3601 CACTACCATG CCCACGTGGA CAGTGGGCCCT GTGTACCCAG AGACCATCTG
3651 TCCCCATAGC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCACCAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGGTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAACTTG
3951 CCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCCTA GCTTGCTCTG
4001 CCCATTCCCTC CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCCCTCCCTC CCCTGTCAAG CCAGTTCTTT CCTCTTCAGG TGGGTGTTCT
4101 GGCCAGCCCC CTTCCCATCC CCAAGGAGCC CTTCAGCGCG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCCC TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CCTGGGCAGC TTATCCTGCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCCT ACAGCTACAT GACAGTGCTG
4451 TTTTATTGTA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTGTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTCACCC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCACCCT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGGAAA CCTGCGTGTC AAGCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGGATTGCCA ACAACTGGAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCAA CAGGAGATGG CCCGCCCTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCTA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAGGTCTT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCTCTT ATTATGTTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAGG GG

```

BLAST Results

Entry HSU64252 from database EMBL:
 Human STS sequence NOTI-225.
 Score = 959, P = 1.2e-36, identities = 195/199

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118
 Category: questionable ORF
 Classification: no clue

```

1 LPLVYALMVP LLSASTLGTL ASDLESVQLC PTATOLGKRS PSVGWGSRRR
51 KAEFGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGDPRPAA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 1

No Alert BLASTP hits found

Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191
Category: putative protein
Classification: no clue

1 MAAAVTQGR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAAARCAPF P

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1
Length = 1,298

HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 36/103 (34%), Positives = 44/103 (42%)

Query: 87 GDESSDPGPQHRAQGGPGPEPTLGLPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
G + PGP G GP P P T+ G S R P PA S P GP +P
Sbjct: 726 GRKRKSPGPAPPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARS GAPAAARCAP 189
AAP AA ++R P+ GP LG W + P+ AP
Sbjct: 783 PAAPRAAAQARPRPVAVSRRAEGPDPLGG-WRRQPPGPSHTAAP 827

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 8/21 (38%), Positives = 9/21 (42%)

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48
DH + A G G AP P
Sbjct: 212 DHAREARAVGRGPSSAAPAAP 232

Pedant information for DKFZphut1_21d15, frame 1

Report for DKFZphut1_21d15.1

[LENGTH] 117
[MW] 11797.32
[pI] 10.68
[KW] Irregular
[KW] SIGNAL PEPTIDE 22
[KW] LOW_COMPLEXITY 38.46 %

SEQ LPLVYALMVPLLSASTLGTLASDLESVQLCPTATQLGKRSPSVGWSRRRKAEPGADAGG
SEGXXX.....
PRD cccccccccccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccc
SEQ SGRAQHPQAPSPSDRGARGPGGRCPCGDCAAARAPPRPLPWARARPGCHGGSGGDRPAA
SEGXXX.....
PRD cc

(No Prosite data available for DKFZphut1_21d15.1)

(No Pfam data available for DKFZphut1_21d15.1)

Pedant information for DKFZphut1_21d15, frame 2

Report for DKFZphut1_21d15.2

```
{LENGTH}      191
{MW}           19916.88
{pI}           10.43
{KW}           TRANSMEMBRANE 1
{KW}           LOW_COMPLEXITY 29.84 %

SEQ  MAAAAVTGQRPETAAAEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG  .....
PRD  cccceeeccccchhhhhhhhhccccccchhhhhhhccccccccccccccccccccchhhh
MEM  .....

SEQ  FLVLMVFVHLYLGNVLALLLVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeeee
MEM  ....MMMMMMMMMMMMMMMM.....

SEQ  TSLPRRAPGPAPRLSPRGPAALPGHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....xxxx
PRD  eccccccccccccccccccccccccccccccccchhhhhhhccccceccccccccchhhhhc
MEM  .....

SEQ  GAPAAARCAPPF
SEG  xxxxxxxxxx..
PRD  cccccccccc
MEM  .....
```

(No Prosite data available for DKFZphut1_21d15.2)

(No Pfam data available for DKFZphut1_21d15.2)

DKFZphutel_22d2

group: signal transduction

DKFZphutel_22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp

Poly A stretch at pos. 3230, no polyadenylation signal found

```

1 CTCTCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
151 CCAGAAAGAGG TTCCTCCCCG GGCAGAAGAA ATCACCATTG CAGCTGATGT
201 CACCCACAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCTCTGT
301 ATAGTGATATG CCGTTAACAA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAATATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTTCAGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGACTGCCC AGAGGAGAAG
601 GAGATGAAC CAGCTTGTAT AAAAGCCCTT ACTCGTATAT TTAATAATATC
651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAAGTC AACTTCTTTC
701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAGCTCTT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCCCTGAAA GGTTTTCTCT TTTTACACAC ACTTTTATAT CAGAGAGGGA
851 GACACGAAC TACTTGGACT GTGCTTCGAC GATTGGTTA TGATGATGAC
901 CTGGATTGTA CACCTGAATA TTTGTTCCCC CTGCTGAAAA TACCTCTGTA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTGGAT AGAGACTGTG CTTTGTCAAC TGATGAGCTT
1051 AAGATTTAT TTAAGTTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCAGTG GACGCTCAGC ACTTATTTAG ATGTACAGCG GTGCCCTGGAA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTGAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAACCTCAAG AAATGTGTTT AGATGTAATG TAATTGGAGT GAAAACTGT
1351 GGGAAAAGTG GAGTTCTTCA GGCTCTTCTT GGAAGAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTGTT TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTGTT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAT CCTTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTACCTTA CTGATTCTGT
1701 CAGGAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGATGGTT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAAT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGTAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCAC CCAGCATTA AAAATAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTTT
2201 CTAATAATTT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCACTTGT
2251 GATATGTCTT TTTAAGTGC TGTAAGAGT AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTACG AGCCACAAT TTCAATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAATTTG CATGTTCTT AATGGCATTC
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTG TTTTGTAAAC AAATAGTTTT
2501 TCAGGCAGTG CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG

```

```

2601 AGTAATAATC AACATTGCT ATAAACCAAG AATGACATT TTCAAAAAGG
2651 TGTGATTG TACAGATTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTGA TGCAACAAAC CCTTGAATT CTATTTGTAT
2751 TCGAAGACAA GTCATTCTTA TTATTATAGA ATAACCAAAA CCTTATTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTCTGTGGG ATTTGTGTA TATTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTA ATCTTGTGAT TACAGTAAC GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAATTCTG TTGTGTGTGT
3201 GTGTGTATA TGATAAATGT ACATACATGG AAAAAAAAA AAAAAAA

```

BLAST Results

Entry AC004527 from database EMBL:
 *** SEQUENCING IN PROGRESS *** NF1-related locus, Direct Submission;
 HTGS phase 1, 10 unordered pieces.
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HS148355 from database EMBL:
 human STS SHGC-31220.
 Score = 1826, P = 7.5e-78, identities = 388/406

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 64 bp to 1803 bp; peptide length: 580
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSI IMSLVSEEF EEVPPRAEEI TIPADVTPE
51 VPTIIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPINNQYTE IETCVCESAK
151 NLKNISELFY YAKAVLHPT GLYCPPEKE MKPACIKALT RIFKISDQDN
201 DGTLNDALN FQRICFNTP LAPQALDVK NVVRKHISDG VADSGTLKG
251 FLFLHTLFIQ RGRHETWTV LRRFGYDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHLDR DCALSPDELK DLFKVPYIP WGPDVNNTVC
351 TNERGWITYQ GFSLQWTLTT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDLQ KKQTRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYAIN TVYVYQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFYECARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPQAFTCN TADAPSKDIF VKLTTMAMP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22d2, frame 1

TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320_4 gene: "SPCC320.04c"; product: "hypothetical protein";
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11.
 Length = 625

HSPs:

Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138
Identities = 263/582 (45%), Positives = 380/582 (65%)

Query: 4 DVRILLVGEPRVGKTSLSIMSLVSEEFPEEVPRAEEITIPADVTPERVPHIVDYSEAEQ 63
DVRIL+L+G+ GKTSLS+MSL+ +E+ + VP R + + IPADVTP E V T IVD S E+
Sbjct: 9 DVRIVLIGDEGCGKTSLSVMSLLEDEWVDVAVPRRLDRVLIPADVTPENVTTISIVDLISKEE 68

Query: 64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
+ + EI QANVIC+VY+V ++ +D + ++W+PLI + + P+ILVGNKSD
Sbjct: 69 DENWIVSEIRQANVICVYVSVTDESTVDGIQTKWLPILIRQSFGEYHETPVILVGNKSDGT 128

Query: 124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKEMKP 183
++ + ILPIM TE+ETCVECSA+ +KN+SE+FYAQKAV++PT PLY + K++
Sbjct: 129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYAQKAVIYPTRPPLYDADTKQLTD 187

Query: 184 ACIKALTRIFKISDQNDGTLNDAELNFFQRICFNTPLAPQAEDEVKNVVRKHISDGVAD 243
KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+
Sbjct: 188 RARKALIRVFKICDRDNDGYLSOTELNDFQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query: 244 SGLTLKGFLFLHTLFIQRGRHETTTWTLRRFGYDDDLDTPEYLFPLLKIPPDCTTELNH 303
L L GFL+LH LFI+RGRHETTW VLR+FGY+ L L+ +YL+P + IP C+TEL+
Sbjct: 248 DSLMLAGFLYLHLLFIERGRHETTWAVLRKFGYETSLKLSYDLYPRITIPVGCSTELSP 307

Query: 304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQGFL 363
F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++
Sbjct: 308 EGVQFVSALFEKYDEDKDGLSPSELQNLFSVCPVPVITKDNILAELETNQRGWLTYNNGYM 367

Query: 364 SQWTLTLYLDVQRCLEYLGYLSILTEQESQAS----AVTTRDKKIDLQKKQTQRNVF 419
+ W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF
Sbjct: 368 AYNNMTTLINLTQTFEQLAYLGFVGRSGPGRAGNTLDSIRVTRERKKDLENHGTDRKVF 427

Query: 420 RCNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLDHI--- 476
+C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLL ++
Sbjct: 428 QCLVVGAKDAGKTVMQSLAGRGMAADVAGIQRH--SPFVINRVRVKESKYLLREVDVL 486

Query: 477 SESEFLTEAEIICDVVCLVYDVSNPKSFHEYCARIFKQHFMDSRIPCLIVAASDLHEVKQ 536
S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q
Sbjct: 487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQYFYRTKTPCVMIATKVEREEVDQ 546

Query: 537 EYSISPTDFCRKHKMPPPPQAFTCNTADAPSKDIFVKLTMMAMP 580
+ + P +FCR+ ++P P F+ S IF +L MA+YP
Sbjct: 547 RWEVPPEEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

Pedant information for DKFZphutel_22d2, frame 1

Report for DKFZphutel_22d2.1

[LENGTH] 580
[MW] 66541.61
[pI] 5.56
[HOMOL] TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07

```

[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
1e-07
[FUNCAT]      08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT]      10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT]      09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 9e-04
[BLOCKS]      BL00410A Dynamin family proteins
[SCOP]         dlp1k_ 3.25.1.3.1 CH-p21 Ras protein [human (Homo sapiens)] 2e-42
[SCOP]         d1guua_ 3.25.1.3.10 RapiA [Human (Homo sapiens)] 5e-59
[PIRKW]        transmembrane protein 1e-79
[PIRKW]        membrane trafficking 2e-06
[PIRKW]        acetylated amino end 3e-09
[PIRKW]        prenylated cysteine 3e-09
[PIRKW]        signal transduction 1e-07
[PIRKW]        transforming protein 3e-09
[PIRKW]        immediate-early protein 8e-06
[PIRKW]        alternative splicing 4e-08
[PIRKW]        P-loop 1e-10
[PIRKW]        lipoprotein 7e-10
[PIRKW]        proto-oncogene 3e-09
[PIRKW]        methylated carboxyl end 3e-09
[PIRKW]        membrane protein 3e-09
[PIRKW]        GTP binding 1e-10
[PIRKW]        thiolester bond 7e-10
[SUPFAM]       ras transforming protein 1e-10
[PROSITE]      ATP_GTP_A 2
[PROSITE]      MYRISTYL 3
[PROSITE]      EF_HAND 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 14
[PROSITE]      TYR_PHOSPHO_SITE 4
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         Ras family (contains ATP/GTP binding P-loop)
[KW]           Irregular
[KW]           3D

```

```

SEQ      MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEITIPADVTPERVPTHIVDYSE
1jai-    ...EEEEEEEEETTTCHHHHHHHHHHCCCCCCCCCEEEEEETEEEEEEEEEECCC
SEQ      AEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLINERTOKDSRLPLILVGNKS
1jai-    CGGGHHHHHHHHHTTEEEEEETTHHHHHHHH-HHHHHHHHHHCTTT-TCEEEEEETT
SEQ      DLVEYSSMETILPIMNQYTEIETCVCESAKNLKNISELFYYAQAVLHPTGPLYCPREEKE
1jai-    TTTTTTTHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....
SEQ      MKPACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALVDKNNVVRKHISDG
1jai-    .....
SEQ      VADSGTLTKGFLFLHTLFIQGRHETTWTVLRRFGYDDDLDTPEYLFPLLKIPDCTTE
1jai-    .....
SEQ      LNHHAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQ
1jai-    .....
SEQ      GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
1jai-    .....
SEQ      CNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLHDISESE
1jai-    .....
SEQ      FLTEAEIICDVVLVDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-    .....
SEQ      SPTDFCRKHKMPPQAFCTCNTADAPSKDIFVKLTMMAMP
1jai-    .....

```

Prosite for DKFZphut1_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

Pfam for DKFZphut1_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGktIK		
	++L+G+ VGK++L ++ EF+EE +P ++ T ++ +++		
Query	6	RILLVGEPRVGKTSLSIMSLVSEEFPEE-VPPR-AEEITIPADVTPERVP	52
HMM	LQIWDTAGQERYRsmRPMYYRGAMGFMVYDITNRqSFENIr.NWweEIr		
	I D E+ + + +A+++ +VY+++N+ S +++++ +W++ I+		
Query	53	THIVDYSEAEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKT		
	+ D+D+ P +LVGNK+DL + ++T + +E+SAK+		
Query	103	ERTDKDSRLPLILVGNKSDLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
	N+ E F+ + +++L + . +++ +++++ + C+		
Query	152	LKNISELFYYAQKAVLHPT-----GLYCPEEKEMK-PACI--	186

DKFZphute1_22e12

group: signal transduction

DKFZphute1_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits
cornichon is required for signal transduction in the EGF-receptor
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp
Poly A stretch at pos. 499, no polyadenylation signal found

```

1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTGCGCGC
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACCTGGT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTTGAA
451 CAAAAAACTA TTTTGCTGT ATTTTACCA TATAAGTAT TAAAAAACA
501 TGAATAAAAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

95300228:
cornichon and the EGF receptor signaling process are necessary for both
anterior-posterior
and dorsal-ventral pattern formation in Drosophila.

Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92
Category: strong similarity to known protein

```

1  MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKQVIP
51  ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398.1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (*Saccharomyces cerevisiae*)
Length = 138

HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCSKLNKWWIPELIGHTIVTV 60
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++
Sbjct: 1 MGAWLFILAVVVCINLFGQVHFTILYADLEADYINPIELCSKVNLITPEAALHGALS 60

Query: 61 LLLMSLHWFIFLLNLPVATWNIYRM 85
L L++ +WF+FLLNLPV +N+ ++
Sbjct: 61 LFLNGYWFVFLNLPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90
+YRMI+ALI
Sbjct: 123 LYRMIMALI 131

Pedant information for DKFZphutel_22e12, frame 1

Report for DKFZphutel_22e12.1

[LENGTH] 92
[MW] 10614.98
[pI] 5.04
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (*Saccharomyces cerevisiae*)
5e-14
[FUNCAT] 03.04 budding, cell polarity and filament formation [*S. cerevisiae*, YGL054c]
2e-15
[PIRKW] transmembrane protein 2e-11
[PROSITE] CK2_PHOSPHO_SITE 3
[KW] SIGNAL PEPTIDE 33
[KW] TRANSMEMBRANE 2

SEQ MEAVVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCSKLNKWWIPELIGHTIVTV
PRD ccchhhhhhhhhhhhhhhhhhhheeeccccccccccccccccceehhhhhhhhhhh
MEMMMMMMMMMMM

SEQ LLLMSLHWFIFLLNLPVATWNIYRMILALIND
PRD hhhhhhhheeeccccchhhhhhhhhhhccccc
MEM MMMMMMMMMMMMMMMMMMMM..MMMMMM....

Prosite for DKFZphutel_22e12.1

PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphutel_22e12.1)

DKFZphut1_22n2

group: uterus derived

DKFZphut1_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```
1 ACAACAGGCT GGTGCTTGG CGTGGAATCC TAAAGTGCC TGGCTTGGAG
51 ACTGGAGTGA GACCCAGGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTGAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAA GACTCCGAAG AGGTTGCAAG TATTTATCT
201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAACTG ATTCTGATTC
401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTG CCAGTTTCTG CTGAAATTAA GGAACCTTTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAACT
551 GAAGCCCTTC ATTCTGATT TTATCCCAGC GTCGGGGAT ATTGATGCAT
601 TCTTAAAGGT CCCACGTCCT GATGAAAGC CTGACAACTT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCACGG TGCTCTCACT
701 CTGGTTAACA GAGATTCTA AGCAGACAA CATCACAAA CATATGAAAG
751 TAAAAAGCCT AGAAGATGCA GAAAAGAATC CCAAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAATCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAAGCTG GAGACATGGA
1151 GACATTAAAC TTCAGCTGAG ACACCTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCC CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCCTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCTGTCT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTTCATTCG AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

BLAST Results

Entry HS188252 from database EMBL:
human STS WI-12265.
Score = 2554, P = 4.1e-109, identities = 556/587

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304
Category: putative protein


```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDD
51 DDDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GVIDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHKVKSLE DAEKNPKAID TWIESISELH RSKPPATVHY
201 TRPMPDIDTL MQEWSPEFEE LLGKVSLEPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFALAE GKKAFTPSSN STSQAGDMET
301 LTFS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae), N = 1,
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae)
Length = 562

HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:   3  DNSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEEHGAPLEG 62
          +  DE EEE++ E++ T          +++DDDDDDDD + D D DDD++E A  G
Sbjct:  497 EEDDDDEEEEDDDEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

Query:   63 AYD 65
          D
Sbjct:  557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:   4  NSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEE 55
          N+ +E ++E +E + T + + N+DDDDDDDD + D D DDD++
Sbjct:  494 NNEEDDDDEEEEDDDEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphutel_22n2, frame 3

Report for DKFZphutel_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION      1
[PROSITE]      CAMP_PHOSPHO_SITE  2
[PROSITE]      CK2_PHOSPHO_SITE  10
[PROSITE]      PKC_PHOSPHO_SITE   1
[PROSITE]      ASN_GLYCOSYLATION  3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY    11.84 %

```

```

SEQ  MADNSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEEHGAPL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cccccccchhhhhchhhhhccccccccccccccccccccccccccccccccccccccccc

SEQ  EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGVIDAFLKVP
SEG  .....
PRD  cccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccc

SEQ  RPDGKPDNLGLLVLDEPSTKQSDPTVLSLWLTENSKQHNIQHKVKSLEDAEKNPKAID
SEG  .....
PRD  cccccccccceeeccccccccccccchhhhhccccccccccccccccchhhhhhhccccccch

SEQ  TWIESISELHRSKPPATVHYTRPMPDIDTLMQEWSPEFEE LLGKVSLEPTAEIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhhhccccccccceeeccccccccchhhhhccccchhhhhccccccccccccchhhhhhh

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFALAEKGKAFTPSSNSTSQAGDMET
SEG  .....

```

Prosite for DKF2phutel_22n2.3			
PS00001	4->8	ASN_GLYCOSYLATION	PD0C00001
PS00001	159->163	ASN_GLYCOSYLATION	PD0C00001
PS00001	290->294	ASN_GLYCOSYLATION	PD0C00001
PS00004	17->21	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	18->22	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	138->141	PKC_PHOSPHO_SITE	PD0C00005
PS00006	5->9	CK2_PHOSPHO_SITE	PD0C00006
PS00006	30->34	CK2_PHOSPHO_SITE	PD0C00006
PS00006	43->47	CK2_PHOSPHO_SITE	PD0C00006
PS00006	45->49	CK2_PHOSPHO_SITE	PD0C00006
PS00006	47->51	CK2_PHOSPHO_SITE	PD0C00006
PS00006	49->53	CK2_PHOSPHO_SITE	PD0C00006
PS00006	168->172	CK2_PHOSPHO_SITE	PD0C00006
PS00006	181->185	CK2_PHOSPHO_SITE	PD0C00006
PS00006	185->189	CK2_PHOSPHO_SITE	PD0C00006
PS00006	235->239	CK2_PHOSPHO_SITE	PD0C00006
PS00009	280->284	AMIDATION	PD0C00009

520

DKFZphut1_22o2

group: uterus derived

DKFZphut1_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="11p15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCCGCG GTCAACAAGG GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCCTCAGGC
151 CCGCCCCCTT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCGCG CCCCGTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGCCAGGG CGGGGCCGCG GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCGGTCCCGC CTTCGCCGCG GCGCCATGGA GCCCGGGCG GTTGCAAGAAG
351 CCGTGAGAGC GGGTGAGGAG GATGTGATTA TGGAAGCTCT GCGGTCATAC
401 AACCAAGGAG ACTCCAGAG CTTCACGTTT GATGATGCCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCCCA CCGTGTATC TGCTGTCAGA GTGTCCGAAT CCTGTCCCGG
551 GACCGCAACT GCCTGGACCC GTTCACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCCTGCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCC GCCTAGTGGT
751 GAAGCTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCACCG
801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGCTAAC GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCCTGTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGAACCCCC
951 CACCCACGCT CCTTCCTTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGACGCT GCCCTTTACC GACACCTGGG GACCTTCTCT CGGCACTGTG
1101 TGATGATGCG TACTGTGGA GACCGCAGC AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGAACTT GCCCTCAAG TGTCTGGATG TTCTCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGGAGTG AATATGGATG
1251 TGATTCGTGC CCTCCTCATC TTCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAAG AGAGTGTAGC TCCCGTGTGC AGCGTGTGTA CTGAATGTGC
1351 CCGGATGCAC CGCCAGCCA GGAAGTTCTT GAAGGCCAGG GGATGGCCAC
1401 CTCGCCAGGT GCTGCCCCCT CTGCGGGATG TGAGGACACG GCCTGAGGTT
1451 GGGGAGATGC TCGGGAACAA GCTTGTCCGC CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTTCTT GTTGTCTCTG TGCTCTGAGA
1551 GTGTGCCCGG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCCTCAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCTGTG GACCGGGAGG GTGGAGGAGA AGCCGCCTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCAGGAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCGGGGTGCA TCTTACGTCC CTGCAGGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTCTGTGCC CCCATCAGGA CTGGTGTGTC TTCCAGAGAC TTCTTGGGG
2001 TTGCAACCTG GGAAGCCAC ATCCCACTGG ATCCACACCC GCCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CCGTTCTGTT CATGATTTCG
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTCAG AAGCTGTGTC CTGAGGGCAG GGCCTACCTG
2251 GGCACACAGA AGAGCATATG GGAGGCGAGG GGTTCGGGTG TGGGTGCACA
2301 CAAAGCAAGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACCTCCCA GGGAGAAAC CTGTGAGAAC TTTCATACG
2401 AGTATATCAG AACACACCTT TCCAAGGTAT GTATGCTCTG TTGTCTCTGT
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCCTTGG
2501 TCTCGTTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAACCTTCC TTCTACTGCT TTTAGACTAC ACTTACAACA
2601 AGGAAAATGC CCCTCGTGTG ACCATAGATT GAGATTATTA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```

BLAST Results

Entry AF015416 from database EMBL:
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:
human STS SHGC-15914.
Score = 1143, P = 9.0e-46, identities = 245/255

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEH SFTFDDAQQ EDRKRLAELL
51 VSVLEQGLPP SHRVIWLSQV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSPVESAD MDVVLES LKLCNLVLSSPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLFL LTALRTDVRQ QLFQELKGVR LLTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCVI IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRMLMTHL DTDVKRVAE
401 FLFVLCSSEV PRFIKYTYG NAAGLLAARG LMAGGRPEGQ YSEDEDTDT
451 EYKEAKASIN PVTGRVEEKP PNPMEGMTTEE QKEHEAMKLV TMFDKLSRN
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDDP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_22o2, frame 2

TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7.
Length = 362

HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query: 215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRHCVM IATAGDRTEEFHG 273
      SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct: 12 SQDNEMVLTEILRLFLPISKRSYLKEEDEQKILL-----LVIEIWASSLNNPNPSPLRW 65

Query: 274 HAVN-LLG-NLPLKCLDVLLTLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
      HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct: 66 HATNALLSFNLQLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query: 328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVRTRP-EVGEMLRNKLVR 386
      + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct: 122 QNTLPPILAILLSLLSFFNIKQNL-----SMLLFPTNDRKQSLQKGKSFRCLLRL 173

Query: 387 MT-HLDTDVKRVAEFLFVLCSSEVPRFIKYTYGNAAGLLAARGLMAGGRPEGQYS--- 442
      +T + + A L LC + + G G A G+ M P + +
Sbjct: 174 LTIPIVEPIGTYIASLLNELCDGDSQQIARIFGAGYAMGISQHSETMPFPSLSKAASPV 233

Query: 443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMTTEEQKEHEAMKLVMTFDKLSRN 499
      + + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct: 234 FQKNSRGQENTEENLAIDPITGSMCTNRNKSQRLE-LSQEEKEREERLFYLFQRLEKN 292

```

```

Query:      500 RVIQ 503
           IQ
Sbjct:      293 STIQ 296

```

Pedant information for DKFZphutel_22o2, frame 2

Report for DKFZphute1_22o2.2

```
[LENGTH]      537
[MW]           60372.53
[pI]           5.20
[BLOCKS]       BL00415L Synapsins proteins
[PROSITE]      MYRISTYL          4
[PROSITE]      CK2_PHOSPHO_SITE    13
[PROSITE]      PKC_PHOSPHO_SITE   10
[PROSITE]      ASN_GLYCOSYLATION   1
[KW]           Ali_Alpha
[KW]           LOW_COMPLEXITY      9.50 %
```

```
SEQ MEPRVAEAVETGEEDVIMEALSRYSNQEHSSQSFTFDDAQQEDRKRLAELLVSVLQEGLPP
SEG .....
PRD ccchhhhhhccccchhhhhhccccccceecchhhhhhcccccccccccccccc
```

```
SEQ SHRWIVLQSVRIISRDRNCLDPFTSRSQLALACYADISVSEGSVPESADMVVLES LKC
SEG .....
PRD ceeeeeeccccccccccccccccccccchhhhhhccccceeeccccccccchhhhhhcccc
```

```
SEQ LCNLVLSSPVAQM LA AEARLVKLTERTVG LYRERSFPHDVQFFDLRLFLLTALRTDVRQ
SEG .....xxxxxxxxxxxxxx
PRD hhhcccchhhhhhccccccccccccccccccccchhhhhhcccccccccccccccc
```

```
SEQ QLFQELKGVRLLTDTLELTGVTPEGNPPPTLLPSQETERAMEILKVLFNITLDSIKGEV
SEG .....
PRD hhhhhcchhhhhhccccccccccccccccchhhhhhcccccccccccccccc
```

```
SEQ DEEDAALYRHGLTLLRHCVMIATAGDRTEEFHGHAVNLLGNLP LKCLDVLLTLEPHGDST
SEG .....
PRD hhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccceeeeeecccccccc
```

```
SEQ EFMGVNDM VIRALLIFLEKRLHKTHRLKESVAPVLSVLTECARMH RPARKFLKAQG WPPP
SEG .....
PRD eeehhhhhhhhhhhhhhhhhhccccceeehhhhhhhchhhhhhcccccccc
```

```
SEQ QVLPLRDVRTREPVGEMLRNKLVR LMT HLD TDVKRVAAEFVLVCSESVP RFIKYTG YG
SEG .....xxx
PRD cccccccccchhhhhhccccccccccccccccchhhhhhccccccccceeecccccccc
```

```
SEQ NAAGLLAARGLMAGGRPEGQYSEDEDTDDEYKEAKASINPVTGRVEEK PPNPMEGMTEE
SEG xxxxxxxxxxxxxxxx.....xxxxxxxxxx
PRD chhhhhhccccccccccccccccccccccccchhhhhhccccccccceeeccccccccchhh
```

```
SEQ QKEHEAMKLVTMFDKLSRRNRVIQPMGMSPRGH LTS LDQAMCETMEQQLS DSD PSDPD
SEG .....xxxxxxxxxx
PRD hhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhcccccccccccccccc
```

Prosite for DKFZphutel 22o2.2

PS000001	230->234	ASN_GLYCOSYLATION	PDOC000001
PS000005	61->64	PKC_PHOSPHO_SITE	PDOC000005
PS000005	69->72	PKC_PHOSPHO_SITE	PDOC000005
PS000005	84->87	PKC_PHOSPHO_SITE	PDOC000005
PS000005	117->120	PKC_PHOSPHO_SITE	PDOC000005
PS000005	145->148	PKC_PHOSPHO_SITE	PDOC000005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000005	324->327	PKC_PHOSPHO_SITE	PDOC000005
PS000005	463->466	PKC_PHOSPHO_SITE	PDOC000005
PS000005	508->511	PKC_PHOSPHO_SITE	PDOC000005
PS000006	12->16	CK2_PHOSPHO_SITE	PDOC000006
PS000006	34->38	CK2_PHOSPHO_SITE	PDOC000006
PS000006	52->56	CK2_PHOSPHO_SITE	PDOC000006
PS000006	99->103	CK2_PHOSPHO_SITE	PDOC000006
PS000006	104->108	CK2_PHOSPHO_SITE	PDOC000006
PS000006	263->267	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006

WO 01/12659

PCT/IB00/01496

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphute1_22o2.2)

DKFZphute1_23e13

group: metabolism

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnological production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1 GGTATTATTA GCTCCTGGCT CCGCTCTAGA CCTCAGCGGT TCTGGCTGCC
51 AGCCTGGGCA GCCTGGGAAG CCTGGGAGGA CGGTGGCTTG CCGGTCTGTC
101 GTGAGGCACT GCGGACGGGG ACCCTCTGGG ATTCTGCTGG ATCTGCCCCG
151 GGGGTACCTT TTGGGGGCTG GGACCCCACT CGAGGGGACA CAACCGTCCC
201 TGGCAGTGGT TGGTTCTGCT TCTCCCTGCA GAAAAGCAGC ATTTTCGGAA
251 GCTGAAGAAT AAGCTAGCCC AGCCACACCA CCTTGTGTG TGACCTTGGG
301 CAGGTGGTTC TGCTCTCTG AGCCTCTGTT TCTCTCTGAG CTGAGCAGCC
351 ACCATGGCTG ACGGTCAGAT GCCCTTCTCC TGCCACTACC CAAGCCGCTT
401 GCGCCGAGAC CCCTTCGGGG ACTCTCCCTT CTCTCTCGC CTGCTGGATG
451 ATGGCTTTGG CATGGACCCC TTCCACAGAG ACTTGACAGC CTCTTGGCCC
501 GACTGGGCTC TGCTCTGCTT CTCTCCGCC TGGCCAGGCA CCCTAAGGTC
551 GGGCATGGTG CCCCAGGGCC CCACTGCCAC CGCCAGGTTT GGGGTGCCTG
601 CCGAGGGGAG GACCCCCCA CCCTTCCCTG GGGAGCCCTG GAAAGTGTGT
651 GTGAATGTGC ACAGCTTCAA GCCAGAGGAG TTGATGGTGA AGACCAAGA
701 TGGATACGTG GAGGTGTCTG GCAACATGA AGAGAAACAG CAAGAAGGTG
751 GCATTGTTTC TAAGAACTTC ACAAGAAAA TCCAGCTTCC TGCAAGGTG
801 GATCCTGTGA CAGTATTTGC CTCCTTTCC CCAGAGGGTC TGCTGATCAT
851 CGAAGCTCCC CAGGTCCCTC CTTACTCAAC ATTTGGAGAG AGCAGTTTCA
901 ACAACGAGCT TCCCCAGGAC AGCCAGGAAG TCACCTGTAC CTGAGATGCC
951 AGTACTGGCC CATCTTGTG TTGTCCCCAA CCCTAGGGCT TCTCTGATTC
1001 CAGGATACAT TACTTTAGCT GAACTCAGAT TTAGTGCAAG TAAATGTTA
1051 GAGGGTGCGG GGGTGAGGAC TGACCACAGA TTCCCTGGAT AGTGTAGTGG
1101 TAGATTCTTC CACAGGATAG CGCAATTGGC AAATCATGCT TGGTTGTGTT
1151 AGGCCAAAAT ACTAGTTTTC CTTTCTTTAC CTTTCTATC TTGATGAAAA
1201 TGTGTGCACAT TCTATAGTTG CAAAACACAT AAAAGGGGAC TTAACATTTT
1251 ACGTTGTATC TTACTTGCAG TGAATGCAAG GGTTACTTTT CTCTGGGGAC
1301 CTCCCCCATC ACCCAGGTTT CTACTCTGGG CTCCCGATTC CCATGGCTCC
1351 CAAACCATGC CGCATGGTTT GGTAAATGAA ACCCAGTAGC TAACCCCACT
1401 GTGCTTCCAC ATGCCTGGCC TAAAATGGGT GATATACAGG TCTTATATCC
1451 CCATATGGAA TTTATCCATC AACCACATAA AAACAAACAG TGCCTTCTGC
1501 CCTCTGCCCA GATGTGTCCA GCACGTTCTC AAAGTTTCCA CATTAGCACT
1551 CCCTAAGGAC GCTGGGAGCC TGTCAGTTTA TGATCTGACC TAGGTCCCCC
1601 CTTTCTCTG TCCCTGTGT TTAAGTCGGG ATTTTACAG AGGGAGCTGT
1651 CTCCAGACAG CTCATCAGG AACCAAGCAA AGGCCAGATA GCCTGACAGA
1701 TAGGCTAGTG GTATTGTGTA TATGGGCGGG ACGTGTGTGT CATTATTATT
1751 TGAGTTATGC TGTGTTTAG GGGTAAATAA CAGTAAATAA TTAATAATAA
1801 TAATAATAAT AATAAAGGAG CTGACGTCTT TAAAAAGAA AAAAAAAAAA
1851 AAAA
```

BLAST Results

Entry HS286348 from database EMBL:
human STS TIGR-A002J47.
Score = 510, P = 1.2e-16, identities = 102/102

Medline entries

95394379:
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:
Physiological and pathological changes in levels of the two
small stress proteins, HSP27 and alpha B crystallin, in rat
hindlimb muscles

Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196
Category: strong similarity to known protein
Prosites motifs: SUBTILASE_ASP (28-39)

1 MADGQMPFSC HYPSRLRRDP FRDSPLSRL LDDGFGMDPF PDDLTA SWPD
51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGE PWKVCV
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P = 4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561_1 product: "heat shock protein HSP27"; Mus musculus
heat shock protein HSP27 internal deletion variant b mRNA, complete
cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog
Length = 209

HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27
Identities = 80/182 (43%), Positives = 102/182 (56%)

Query: 1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRLDDGFGMDPFPPDDLTA SWPDWALPRLSS 58
M + ++PFS PS DPFRD P SRL D FG+ P++ W W S
Sbjct: 1 MTERRVFFSLLRSPSW---DPFRDWYPAHSRLFDQAFGLPRLPEE---WAQWFG---HS 50

Query: 59 AWPCTLRSGMVP---RGPTATARFGVPAEGR--TPPPFG-----EPWKVCNVVHSF 105
WPG +R +P GP A A PA R + G + W+V ++V+ F
Sbjct: 51 GWPGYVRP--IPPAVEGPAAAAAAPAYSRLSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query: 106 KPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLI 165
PEEL VKTKDG VE+GKHEE+Q E G +S+ T K LP VDP V +SLSP EG L
Sbjct: 109 APEELTVKTKDGVVEITGKHEERQDEHGYISRLTPKYTLPPGVDP TLVSSSLSP EGT LT 168

Query: 166 IEAPQVPPYSTFGE 179
+EAP P + E
Sbjct: 169 VEAPMPKPATQSAE 182

Pedant information for DKFZphut1_23e13, frame 3

Report for DKFZphut1_23e13.3

[LENGTH] 196
[MW] 21604.37


```

[pI]          5.00
[HOMOL]       PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS]      BL01031C
[PIRKW]       blocked amino end 1e-13
[PIRKW]       acetylated amino end 4e-13
[PIRKW]       phosphoprotein 7e-21
[PIRKW]       glycoprotein 2e-11
[PIRKW]       heat shock 7e-21
[PIRKW]       molecular chaperone 4e-13
[PIRKW]       alternative splicing 1e-19
[PIRKW]       eye lens 6e-14
[PIRKW]       stress-induced protein 7e-21
[SUPFAM]      alpha-crystallin 7e-21
[PROSITE]     SUBTILASE_ASP 1
[PROSITE]     MYRISTYL 2
[PROSITE]     CK2_PHOSPHO_SITE 2
[PROSITE]     PKC_PHOSPHO_SITE 6
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        Heat shock hsp20 proteins
[KW]          All_Beta
[KW]          LOW_COMPLEXITY 7.14 %

```

```

SEQ  MADGQMPFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTAWPDWALPRLSSAW
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PGTLRSGMVPFRGPTATARFGVPAEGRTPPPFPGEPWKVCNVVHSFKPEELMVKT KDGYVE
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VSGKHEEKQEGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLI IEAPQVPPYSTFGES
SEG  .....
PRD  eccchhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFNNELPQDSQEVCTCT
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphut1_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

Pfam for DKFZphut1_23e13.3

```

HMM_NAME      Heat shock hsp20 proteins
HMM            *AMMrpPwDWRE....DpDHFeVrMDMPGFKPEEIKVkvEDNNVLvIeG
               A   P++ R   + ++V++++ FKPEE+ VK+ D+ +++++G
Query         77 ARFGVPAEGR-TPPPFPGEPWKVCNVVHSFKPEELMVKT KDG-YVEVSG 123

HMM            EHEREEEREDdkWWHERIYRHFMRFRrLPENVDPdQIkAsMSdNGVLTi
               +HE E++   + + + F +++LP +VDP + AS+S++G+L I
Query         124 KHE---EKQQ----EGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLI 166

HMM            TVPKpEP*
               ++P ++P
Query         167 EAPQVPP 173

```

DKFZphute1_23g11

group: uterus derived

DKFZphute1_23g11 encodes a novel 256 amino acid protein with similarity to *S.pombe* SPAC31G5.12c and *S. cerevisiae* Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GCGGTCGCGG GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGATGGG CCGGCAGGGC
151 CCGGAGTAAC GGGACGTCGC CGCGGAGCTT CTCCCCCGG ATACAGTGCG
201 GCGCGAGCGG AGGCCGCGGC GCCGCCCTCC GATCTTGAAG AGCCCCGCGT
251 GCGCGGAGCC CGCCCCCGCC TGCGCACCAG CACCGACGCG GAGCGACCCG
301 CCCAGCCAGA CCCGGCCCCG CGCGGCCTGA TCTAACCCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAG ACATGAAGCT
401 ATTGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGTACTC ATGTAAGATG
501 GCAGGAGACG ACAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACTTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAAGCCAA GCGGTGAGG AGGAGGGCCC CCTCAGTGAC
651 AAGTGAGGCC GCAAGACCCT TTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
751 GGGAGGCCAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCAAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCTTC
951 AAGTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCCCAGCT
1251 GCTGGGCCAGA CCTGGCGCTC GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCCTAGC CCTTTGGCTC CATCCTGTGG ATGCCCACTC ACCCTCAGA
1351 CTCTCTGCTG CCAATGCTGTG GCCGGAATTG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGGAGCT GCCCTGCCCA AATGAATGCG CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCAAA
1501 GGTCTGTGGG CCGGAGGCCC CACGAGCAGG CCCAGCAGT CACCGGCTCT
1551 GGTCTTGGGC CGGCCCCGGT GCCCACCCTG ACCCCACCTC CGCCCATTTG
1601 GCGCGCTGCA CTGAGTGTC ATTTGCTGCA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256
Category: similarity to known protein

```

1 MKLLENSFE AINSQTLVET GDAHIIGRIE SYSCKMAGDD KHMFKQFCQE
51 GQPHVLEALS PPQTSGLSPLS RLSKSQGGEE EGPLSDKCSR KTLFYLIATL
101 NESFRPDYDF STARSHEFSR EPSLSWVVNA VNCSLFSAVR EDFKDLKPQL
151 WNAVDEEICL AECDIYSYNP DLSDPFGED GSWSFNYFF YNKRLKRIVF
201 FSCRSISGST YTPSEAGNEL DMELGEEVEE EESRSRGSGA EETSTMEEDR
251 VPVICI

```

BLASTP hits

Entry SPAC31G5.12 from database TREMBL:
 gene: "SPAC31G5.12c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c31G5.
 Score = 272, P = 9.3e-24, identities = 51/127, positives = 80/127

Entry SPD656.1 from database TREMBL:
 product: "ORF N150"; Yeast DNA for bfr2+ protein/pad1+ protein/sks1+
 protein, ORF N313, ORF N150, complete cds, and for ORF N118, partial
 cds.
 Score = 263, P = 8.4e-23, identities = 50/127, positives = 79/127

Entry S50986 from database PIR:
 MAF1 protein - yeast (Saccharomyces cerevisiae) >SWISSPROT:MAF1_YEAST
 MAF1 PROTEIN. >TREMBL:SC19492.1 gene: "MAF1"; product: "Maf1p";
 Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds.
 >TREMBL:SC8119.11 gene: "MAF1p"; product: "Maf1p"; S.cerevisiae
 chromosome IV Cosmid 8119.
 Score = 180, P = 2.3e-17, identities = 43/133, positives = 75/133

Entry AF098499.2 from database TREMBL:
 gene: "C43H8.2"; Caenorhabditis elegans cosmid C43H8.
 Score = 263, P = 9.2e-23, identities = 78/252, positives = 118/252

Alert BLASTP hits for DKFZphut1_23g11, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_23g11, frame 3

Report for DKFZphut1_23g11.3

```

[LENGTH]      256
[MW]           28869.95
[pI]           4.51
[HOMOL]        TREMBL:SPAC31G5.12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR005c]
6e-13
[PROSITE]      MYRISTYL      3
[PROSITE]      CK2_PHOSPHO_SITE      5
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      7.81 %

```

```

SEQ  MKLLENSFEAINSQTLVETGDAHIIGRIESYSCKMAGDDKHMFKQFCQEGQPHVLEALS
SEG  .....
PRD  cccccchhhhhhhhhhhccccceeeccchhhhhccchhhhhhhhhccccceeeccccc

SEQ  PPQTSGLSPLSRLSKSQGGEEEGPLSDKCSRKTLFYLIATLNESEFRPDYDFSTARSHEFSR
SEG  .....
PRD  cccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccc

SEQ  EPSLSWVVNAVNCSLFSAVREDFKDLKPQLWNAVDEEICLAECDIYSYNPDLSDFPFGED
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccecccccccccccc

SEQ  GSWSFNYFFYNKRLKRIVFFSCRSISGSTYTPSEAGNELDMELGEEVEEESRSRGSGA
SEG  .....
PRD  ccceeeceechhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhcccccccc

SEQ  EETSTMEEDRVPVICI
SEG  xx.....
PRD  cccccccccceeeccc

```

Prosites for DKFZphut1_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_23g11.3)

DKFZphutel_24c19

group: transmembrane protein

DKFZphutel_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown

membrane regions: 1

Summary DKFZphutel_24c19 encodes a novel 195 amino acid protein, with no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp

Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
1 ACGAGTCAGC CAAAGATGGC TGC GCCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACCTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAACACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAACAGT CTTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAAGTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTCCT ATTTGCTCC AGACTATGTT TTCAGCATAC CTTGGGTCTG
651 AACATATAA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCAC TATTTTAAAC AAATATGTAA ACAAATAAAT AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195
Category: putative protein

```
1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51 NSLFRRILNV TKARIAAGLP MAGIPFLTDD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGEQYKL LIKALQLSEP GKEIH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1_24c19, frame 2

Report for DKFZphut1_24c19.2

```

[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSLFRRLNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPAGIPFLTDTLTYSRCFVSFPLNTGDLDCETCTITRSGTLGLVIGGLYPVF
PRD  hhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPILLQTMFSAYLGSEQYKL
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....

```

Prosite for DKFZphut1_24c19.2

```

PS00001      11->15  ASN_GLYCOSYLATION      PDOC00001
PS00001      34->38  ASN_GLYCOSYLATION      PDOC00001
PS00001      59->63  ASN_GLYCOSYLATION      PDOC00001
PS00005      18->21  PKC_PHOSPHO_SITE      PDOC00005
PS00005      82->85  PKC_PHOSPHO_SITE      PDOC00005
PS00005     151->154 PKC_PHOSPHO_SITE      PDOC00005
PS00006      13->17  CK2_PHOSPHO_SITE      PDOC00006
PS00008      40->46  MYRISTYL              PDOC00008
PS00008      47->53  MYRISTYL              PDOC00008
PS00008      68->74  MYRISTYL              PDOC00008
PS00008     110->116 MYRISTYL              PDOC00008
PS00008     127->133 MYRISTYL              PDOC00008
PS00008     142->148 MYRISTYL              PDOC00008

```

(No Pfam data available for DKFZphut1_24c19.2)

DKFZphutell_24e11

group: intracellular transport and trafficking

DKFZphutell_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits
potential start at 184,
TRANSMEMBRANE 4
function in the transport of nucleosides and/or nucleoside derivatives
between the cytosol and
the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp

Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
1  ACGCGTCCGG  CAGAAGCTCG  GAGCTCTCGG  GGTATCGAGG  AGGCAGGCC
51  GCGGGCGCAC  GGGCGAGCGG  GCCGGGAGCC  GGAGCGGCGG  AGGAGCCGG
101  ACGACGCGCG  CGGCGGGCTC  CAGGCGAGGC  GGTGACGCT  CCTGAAACT
151  TCGCGCGCGG  CTCGCGCCAC  TCGCGCCGGA  GCGATGAAGA  TGGTCGCGCC
201  CTGGACGCGG  TTCTACTCCA  ACAGCTGCTG  CTTGTGCTGC  CATGTCGCG
251  CCGGCACCAT  CCTGCTCGGC  GTCTGGTATC  TGATCATCAA  TGCTGTGGTA
301  CTGTTGATTT  TATTGAGTGC  CCTGGCTGAT  CCGGATCAGT  ATAACTTTTC
351  AAGTTCTGAA  CTGGGAGGTG  ACTTTGAGTT  CATGGATGAT  GCCAACATGT
401  GCATTGCCAT  TCGGATTTCT  CTTCTCATGA  TCCTGATATG  TGCTATGGCT
451  ACTTACGGAG  CGTACAAGCA  ACGCGCAGCC  TGGATCATCC  CATTCTTCTG
501  TTACCAGATC  TTTGACTTTG  CCCTGAACAT  GTTGGTTGCA  ATCACTGTGC
551  TTATTTATCC  AAATCCATT  CAGGAATACA  TACGGCAACT  GCCTCCTAAT
601  TTTCCCTACA  GAGATGATGT  CATGTCAGTG  AATCCTACCT  GTTTGGTCTT
651  TATTATTCTT  CTGTTTATTA  GCATTATCTT  GACTTTTAAG  GGTACTTTGA
701  TTAGCTGTGT  TTGGAAGTGC  TACCGATACA  TCAATGGTAG  GAACTCCTCT
751  GATGTCTCTG  TTTATGTTAC  CAGCAATGAC  ACTACGGTGC  TGCTACCCCC
801  GTATGATGAT  GCCACTGTGA  ATGGTGTCTG  CAAGGAGCCA  CCGCCACCTT
851  ACGTCTCTGC  CTAAGCCTTC  AAGTGGGCGG  AGCTGAGGGC  AGCAGCTTGA
901  CTTTGCAGAC  ATCTGAGCAA  TAGTTCTGTT  ATTTCACTTT  TGCCATGAGC
951  CTCCTCTGAG  TTGTTTGTG  CTGAATGCT  ACTTTTAA  ATTTAGATGT
1001  TAGATTGAAA  ACTGTAGTTT  TCAACATATG  CTTTGCTAGA  ACACCTGTGAT
1051  AGATTAACTG  TAGAATCTT  CCTGTACGAT  TGGGGATATA  ACGGGCTTCA
1101  CTAACCTTCC  CTAGGCATTG  AAATCTCCC  CAAATCTGAT  GGACCTAGAA
1151  GTCTGCTTTT  GTACCTGCTG  GGCCCCAAG  TTGGGCATTT  TTCTCTCTGT
1201  TCCCTCTCTT  TTGAAAATGT  AAAATAAAAC  CAAAATAGA  CAACTTTTTC
1251  TTCAGCCATT  CCAGCATAGA  GAACAAAACC  TTATGGAAAC  AGGAATGTCA
1301  ATTGTGTAAT  CATTGTTCTA  ATTAGGTAAA  TAGAAGTCCT  TATGTATGTG
1351  TTACAAGAA  TTCCCCACA  ACATCCTTTA  TGAATGAAGT  TCAATGACAG
1401  TTTGTGTTTG  GTGGTAAAG  ATTTCTCCA  TGGCTGAAT  TAAGACCAT
1451  AGAAAGCAC  AGCCGTGGG  AGCAGTGACC  ATCTACTGAC  TGTCTTGTG
1501  GATCTTGTGT  CCAGGGACAT  GGGGTGACAT  GCCTCGTATG  TGTAGAGGG
1551  TGGAAATGG  GTGTTTGGCG  CTGCATGGGA  TCTGGTGCCC  CTCTTCTCT
1601  GGGATTCAT  CCCCACCCAG  GGCCCGCTTT  TACTAAGTGT  TCTGCCCTAG
1651  ATTGGTTCAA  GGAGGTCATC  CACTGACTT  TATCAAGTGG  AATTGGGATA
1701  TATTGTATAT  ACTTCTGCCT  AACAACATGG  AAAAGGGTTT  TCTTTTCCCT
1751  GCAAGCTACA  TCCTACTGCT  TTGAACCTCC  AAGTATGTCT  AGTCACCTTT
1801  TAAAATGTAA  ACATTTTCAG  AAAAATGAGG  ATTGCCTTCC  TTGTATGCGC
1851  TTTTACCTT  GACTACCTGA  ATTGCAAGGG  ATTTTATAT  ATTCATATGT
1901  TACAAAGTCA  GCAACTCTCC  TGTGTTTCA  TTATTGAATG  TGCTGTAAT
1951  TAAGTCGTTT  GCAATTAATA  CAAGTTTTC  CCACATCCAA  AAAAAAAAAA
2001  AAAAA
```

BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.
Score = 1629, P = 3.1e-67, identities = 343/354

Medline entries

96199248:
Identification of a novel membrane transporter
associated with intracellular membranes by
phenotypic complementation in the yeast
Saccharomyces cerevisiae.

Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226
Category: strong similarity to known protein

1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP
51 DQYNFSSSEL GGDFFEMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN
151 PTCLVLIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT
201 TVLLPPYDDA TVNGAAKEPP PPYVSA

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24e11, frame 1

SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981_1 product: "E3 protein"; Human retinoic acid-inducible
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108).
Length = 233

HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53
Identities = 102/221 (46%), Positives = 148/221 (66%)

Query: 9 RFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSSELGGDF- 64
RFYS CC CCHVRTGTI+LG WY+++N ++ ++L + P+ N +G +
Sbjct: 13 RFYSTRCCGCCCHVRTGTIILGTWYVMVNLMAILLTVEVTHPNSMPAVNIQYEVIGNYYS 72

Query: 65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123
E M D N C+ A+S+LM +I +M YGA + W+IPFFCY++FDF L+ LVAI+ L
Sbjct: 73 SERMAD-NACVLFASVLMFIISMLVYGAIYQVGLIPFFCYRLDFVLSCLVAISSL 131

Query: 124 IYPNSIQEYIRQLPPNFYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYI 183
Y I+EY+ QLP +FPY+DD++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI
Sbjct: 132 TYLPRIKEYLDQLP-DFFPKDDLALDSSCLLFIVLVFFALFIIKAYLINCWNCYKYI 190

Query: 184 NGRNSSDVLVYVTSN-DTIVLLPPYDDATVNGAAKEPPPPYVSA 226
N RN ++ VY +LP Y+ A V KEPPPPY+ A
Sbjct: 191 NNRNVPEIAVYPAFEAPPQYVLPYEMA-VKMPKEPPPPYLP 233

Pedant information for DKFZphut1_24e11, frame 1

Report for DKFZphut1_24e11.1

[LENGTH] 226
[MW] 25419.11

[pI] 4.65
 [HOMOL] SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).
 5e-40
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 1
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] SIGNAL PEPTIDE 49
 [KW] TRANSMEMBRANE 2
 [KW] LOW_COMPLEXITY 20.80 %

SEQ MKMVAPWTRFYSNSCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSEL
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEM

SEQ GGDFFMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFCYQIFDFALNMLVAI
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEMMM

SEQ TVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCY
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD hhhcc
 MEM MMMMMM.....MM

SEQ RYINGRNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA
 SEG
 PRD ecc
 MEM

Prosite for DKFZphut1_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphut1_24e11.1)

DKF2phutel_24j6

group: cell structure and motility

DKF2phutesl_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CAR1).

The novel protein is very similar to Car1 and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CAR1 A.thaliana T19C21.5

complete cDNA, complete cds, EST hits
potential frame shift at Bp 1241 according to CAR1
but frame shift might be in CAR1 sequence!
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```
1  ACGCGTCCGA GCTGGCTCAG GCGGTCCGCT AGGCTCGGAC GACCTGCTGA
51  GCCTCCCAAA CCGCTTCCAT AAGGCTTTGC CTTTCCAAC TCAGCTACAG
101 TGTTAGCTAA GTTTGGAAAG AAGGAAAAAA GAAATCCCT GGGCCCCCTT
151 TCTTTTGTTC TTTGCCAAAG TCGTCGTTGT AGTCTTTTG CCAAAGGCTG
201 TTTGTGTTTT AGAGGTGCTA TCTCCAGTTC CTGCACTCC TGTTAACAAG
251 CACCTCAGCG AGAGCAGCAG CAGCGATAGC AGCCGCAGAA GAGCCAGCGG
301 GGTCCGCTAG TGTATGACC AGGCGGGGAG ATCACAACCG CCAGAGAGGA
351 TGCTGTGGAT CCTTGGCCGA CTACCTGACC TCTGCAAAAT TCCTTCTCTA
401 CCTTGGTCAT TCTCTCTCTA CTTGGGGAGA TCGGATGTGG CACTTTGCGG
451 TGTCTGTGTT TCTGGTAGAG CTCTATGGAA ACAGCCTCCT TTTGACAGCA
501 GTCTACGGGC TGGTGGTGGC AGGGTCTGTT CTGGTCTGG GAGCCATCAT
551 CGGTAGCTGG GTGGACAAGA ATGCTAGACT TAAAGTGGCC CAGACCTCGC
601 TGGTGGTACA GAATGTTTCA GTCATCCTGT GTGGAATCAT CCTGATGATG
651 GTTTTCTTAC ATAAACATGA GCTTCTGACC ATGTACCATG GATGGGTCTT
701 CACTTCTCTG TATATCCTGA TCATCACTAT TGCAATATAT GCAAAATTGG
751 CCAGTACTGC TACTGCAATC ACAATCAAAA GGGATTGGAT TGTGTTGTTT
801 GCAGGAGAAG ACAGAAGCAA ACTAGCAAAAT ATGAATGCCA CAATACGAAG
851 GATTGACCAG TTAACCAACA TCTTAGCCCC CATGGCTGTT GGCCAGATTA
901 TGACATTGGG CTCCCCAGTC ATCGGCTGTG GCTTTATTTC GGGATGGAAAC
951 TTGGTATCCA TGTGCGTGGA GTACGTCCTG CTCTGGAAGG TTTACCAGAA
1001 AACCCAGCTC CTAGCTGTGA AAGCTGGTCT TAAAGAAGAG GAAACTGAAT
1051 TGAACACAGT GAATTTACAC AAAGATACTG AGCCAAAACC CCTGGAGGGA
1101 ACTCATCTAA TGGGTGTGAA AGACTCTAAC ATCCATGAGC TTGAACATGA
1151 GCAAGAGCCT ACTTGTGCCT CCCAGATGGC TGAGCCCTTC CGTACCTTCC
1201 GAGATGGATG GGTCTCCTAC TACAACCAGC CTGTGTTTCT GGCTGGCATG
1251 GGTCTTGCTT TCCTTTATAT GACTGTCCCTG GGCTTTGACT GCATCACCAC
1301 AGGGTACGCC TACACTCAGG GACTGAGTGG TTCCATCCTC AGTATTTTGA
1351 TGGGAGCATC AGCTATAACT GGAATAATGG GAACGTAGC TTTTACTTGG
1401 CTACGTCGAA AATGTGGTTT GGTTCGGACA GGTCTGATCT CAGGATTGGC
1451 ACAGCTTTCC TGTGTTGATC TGTGTGTGAT CTCTGTATTC ATGCCTGGAA
1501 GCCCCTGGA CTTGTCCGTT TCTCCTTTTG AAGATATCCG ATCAAGGTTT
1551 ATTCAAGGAG AGTCAATTAC ACCTACCAAG ATACCTGAAA TTACAACCTGA
1601 AATATACATG TCTAATGGGT CTAATCTGTC TAATATTGTC CCGGAGACAA
1651 GTCCTGAATC TGTGCCATA ATCTCTGTCA GTCTGCTGTT TGCAGGCGTC
1701 ATTGCTGCTA GAATCGGTCT TTGGTCCTTT GATTTAAGTG TGACACAGTT
1751 CTGCAAGAAA AATGTAATG AATCTGAAAG AGGCATTATA AATGGTGTAC
1801 AGAATCCCAT GAATATCTT CTTGATCTTC TGCATTTCAT CATGGTCATC
1851 CTGGCTCCAA ATCCTGAAGC TTTTGGCTTG CTCGTATTGA TTTCACTCTC
1901 CTTTGTGGCA ATGGGCCACA TTATGTATT CCGATTGTGC CAAAATACTC
1951 TGGGAAACAA GCTCTTTGCT TCGGCTCCTG ATGCAAAAGA AGTTAGGAAG
2001 GAAATCAAG CAAATACATC TGTGTTTGA GACAGTTTAA CTGTTGCTAT
2051 CTTGTTACTA GATTATATAG AGCACATGTG CTTATTTTGT ACTGCAGAAAT
2101 TCCAATAAAT GGCTGGGTGT TTTGCTCTGT TTTTACCACA GCTGTGCCCT
2151 GAGAACATAA AGCTGTTTAG GAAACCTAAG TCAGCAGAAA TTAAGTGATT
2201 AATTTCCCTT ATGTTGAGGC ATGGAAAAAA AATTGGAAAA GAAAACTCA
2251 GTTTAATATC GGAGACTATA ATGATAACAC TGAATCCCTC TATTTCTCAT
2301 GAGTAGATAC AATCTTACGT AAAAGAGTGG TTAGTCACGT GAATTCAGTT
2351 ATCATTGTAC AGATTCTTAT CTGTACTAGA ATTCAGATAT GTCAGTTTTC
2401 TGCAAAACCTC ACTCTTGTTC AAGACTAGCT AATTTATTTT TTTGCATCTT
2451 AGTTATTTT AAAACAAAAT TCTTCAAGTA TGAAGACTAA ATTTTGATAA
2501 CTAATATTAT CCTATTGAT CCTATTGATC TTAAGGTATT TACATGTATG
```

```

2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAAGCACT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAAGTA GATTATACAC ACAATGACTG CATACAGACT
2751 TCAAAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACATCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGATA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCACACA CTTGTGAAGG TTTTGTTTA CAAATCACTT GATTTAACAC
2951 ACTCAGGTAG AATATTTTAA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAGTC AGTTTGCAAC ATGCTGTAC CAAGATGGTA CTTTGCCTTA
3151 ACCGTTTATA TGCACCTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TCTCTTTATC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATTT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA

```

BLAST Results

Entry HS389210 from database EMBL:
human STS SHGC-10164.
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:
human STS WI-16551.
Score = 1193, P = 5.7e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571
Category: strong similarity to known protein

```

1 MTRAGDHNRO RGCCGSLADY LTSAKFLLYL GHSLSTWGDH MWHFAVSVFL
51 VELYGNLALL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKKEL LTMVHGWLVT SCYILITIA NIANLASTAT
151 AITTIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNLLAPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCUEY VLLWKVYQRT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDQWV
301 SYYNQPVFLA GMGLAFLYMT VLGFDCTTG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAE TWLRRKGLV RTGLISGLAQ LSCLILCVIS VFMPSPLDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEE YMSNGSNSAN IVPETSPESV
451 PIISVSLLEA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGKGL
551 FACGPDKEV RKENQANTSV V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24j6, frame 3

TREMBLNEW:U76714.1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683.5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P
= 2.8e-60

TREMBL:AF039046.2 gene: "R09B5.4"; Caenorhabditis elegans cosmid
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714.1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.
Length = 405

HSPs :

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151
Identities = 288/319 (90%), Positives = 297/319 (93%)

Query:	1	MTRAGDHHNRQRCGCSLADYLTSAKFLLYLGHSLSTWGRMMHFAVSVFVLELYGNSLL	60
		MT++ D Q GCGGSLA-YLTSAKFLLYLGHSLSTWGRMMHFAVSVFVLELYGNSLL	
Sbjct:	1	MTKSRDQTHQEGCCGSLANYLTSAKFLLYLGHSLSTWGRMMHFAVSVFVLELYGNSLL	60
Query:	61	TAVYGLVVGAGSVLVLGAIGIDGWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLHKHEL	120
		TAVYGLVVGAGSVLVLGAIGIDGWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLH++EL	
Sbjct:	61	TAVYGLVVGAGSVLVLGAIGIDGWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLHKNEL	120
Query:	121	LTMHYHGWLVTSCYILIITIANIANLASTATAITIQRDWIVVVAGEDRSKLANMNAITIRRI	180
		L MYHGWLVT CYILIITIANIANLASTATAITIQRDWIVVVAGE+RS+LA+MNAITIRRI	
Sbjct:	121	LNMHYHGWLVTVCYILIITIANIANLASTATAITIQRDWIVVVAGENRSALADMNATIRRI	180
Query:	181	DQLTNI LAPMAVGQIMTFGSPVIGCGFISGWNLVSMCV EYLLWKVYQKTPALAVKAGLK	240
		DQLTNI LAPMAVGQIMTFGSPVIGCGFISGWNLVSMCV EY LLWKVYQKTPALAVKA LK	
Sbjct:	181	DQLTNI LAPMAVGQIMTFGSPVIGCGFISGWNLVSMCV EYLLWKVYQKTPALAVKAAKL	240
Query:	241	EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVV	300
		EE+ELKQL KDTEPKPLEGTHLMG KDSNI ELE EQEPTCASQ+AEFPRTFRDGVV	
Sbjct:	241	VEESELKQLTSPKDTTEPKPLEGTHLMGEKDSNI RELEQEPTCASQIAEPFRTFRDGVV	300
Query:	301	SYYNQPVFLAGMGLAF-LY 318	
		SYYNQPVFL G FLY	
Sbjct:	301	SYYNQPVFLGWHGFGPGFLY 319	

Pedant information for DKF2phutel 24j6, frame 3

Report for DKFZphutel1_24j6.3

```

[LENGTH]          571
[MW]               62542.72
[pI]               6.08
[COMOL]            TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator.(CAR1) mRNA, complete cds. 1e-141
[BLOCKS]           BL00341D
[PROSITE]          MYRISTYL      15
[PROSITE]          MITOCH_CARRIER 1
[PROSITE]          CK2_PHOSPHO_SITE      6
[PROSITE]          PROKAR_LIPOPROTEIN    1
[PROSITE]          PKC_PHOSPHO_SITE      4
[PROSITE]          ASN_GLYCOSYLATION     4
[PFAM]             Laminin B (Domain IV)
[KW]               TRANSMEMBRANE 4
[KW]               LOW COMPLEXITY      8.76 %

```

```

SEQ      MTRAGDHNHRQGCCGSLADYLTSAKFLLYLGHLSLSTWGRDMWHFAVSVFLVELYGNLSLL
SEG
PRD      cccccccccccccccccchhhhhhhhhheeeccceeeccchhhhhhhhhheeecccccce
MEM      .....MMMMMMMMMMMMMM

SEQ      TAVYGLVVGASVLVLGAIIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG      .xxxxxxxxxxxxxxxx
PRD      ehhhhhhhhhccceeeccccchhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      LTMYHGWVLTSCYLIITIANINLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI
SEG      .xxxxxxxxxxxxxxxxxxxx
PRD      hhccccchhhhhhhhhhhhhhhhhhhhhhhheeeccceeeccccchhhhhhhhhhhhh
MEM      MMMMMMMM.....

SEQ      DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG
PRD      hhhhhhhccceeeceeeeeeccceeeeeeccchhhhhhhhhhhhhhhhhccchhhhhhhhh
MEM      .....

SEQ      EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVV
SEG
PRD      hhhhhhhhhhhccccccccccccceeeeeeccccccccccccccccccccccccccccce
MEM      .....

SEQ      SYYNQPVFLAGMGLAFLYMTVLGFDCTTGYAYTQGLSGSILSILMGASAITGIMGTAVF
SEG
PRD      eeccceeeccchhhhhhhccccceeeeeeccccceeeeeeccccceeeehhhhhhh

```

[illegible]

Prosites for DKFZphutel1_24j6.3

PS000001	100->104	ASN_GLYCOSYLATION	PDOC000001
PS000001	174->178	ASN_GLYCOSYLATION	PDOC000001
PS000001	434->438	ASN_GLYCOSYLATION	PDOC000001
PS000001	567->571	ASN_GLYCOSYLATION	PDOC000001
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	176->179	PKC_PHOSPHO_SITE	PDOC000005
PS000005	294->297	PKC_PHOSPHO_SITE	PDOC000005
PS000005	487->490	PKC_PHOSPHO_SITE	PDOC000005
PS000006	16->20	CK2_PHOSPHO_SITE	PDOC000006
PS000006	36->40	CK2_PHOSPHO_SITE	PDOC000006
PS000006	294->298	CK2_PHOSPHO_SITE	PDOC000006
PS000006	396->400	CK2_PHOSPHO_SITE	PDOC000006
PS000006	403->407	CK2_PHOSPHO_SITE	PDOC000006
PS000006	445->449	CK2_PHOSPHO_SITE	PDOC000006
PS000008	12->18	MYRISTYL	PDOC000008
PS000008	65->71	MYRISTYL	PDOC000008
PS000008	76->82	MYRISTYL	PDOC000008
PS000008	193->199	MYRISTYL	PDOC000008
PS000008	267->273	MYRISTYL	PDOC000008
PS000008	311->317	MYRISTYL	PDOC000008
PS000008	336->342	MYRISTYL	PDOC000008
PS000008	339->345	MYRISTYL	PDOC000008
PS000008	353->359	MYRISTYL	PDOC000008
PS000008	368->374	MYRISTYL	PDOC000008
PS000008	373->379	MYRISTYL	PDOC000008
PS000008	435->441	MYRISTYL	PDOC000008
PS000008	461->467	MYRISTYL	PDOC000008
PS000008	490->496	MYRISTYL	PDOC000008
PS000008	494->500	MYRISTYL	PDOC000008
PS000113	122->133	PROKAR_LIPOPROTEIN	PDOC000113
PS002115	404->414	MITOCH_CARRIER	PDOC001189

Pfam for DKFZphute1 24j6.3

HMM_NAME	Laminin B (Domain IV)		
HMM	*YWRIPERFLGDQvTsYGGKle*		
	Y+R	+ LG+++ + G + +	
Query	538	YERFAQNTLGKFLACGPDAK	558

DKF2phutel_2h3

group: differentiation/development

DKF2phutel_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits
complete cds according to E25 start at Bp 56
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp

Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```
1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GCGGACGCG CGGGCCGGCG
51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
101 GACAAGGCTG ACAAGGCGTC GCGTCGGCC CCTGCGCCGG CCTCGGCCAC
151 CGAGATCCTG CTGACGCCGG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
201 CCAAGAGGGG GAGCTCAGTG GCGGCGGTGT GCTACCTGTC GATGGGCATG
251 GTCTGTCTGC TCATGGGCCT CGTGTTCGCC TCTGTCTACA TCTACAGATA
301 CTTCTTTCTT GCACAGCTGG CCGAGATAA CTTCTCCGC TGTGGTGTGC
351 TGTATGAGGA CTCCTGTCC TCCAGGTCC GGACTCAGAT GGAGCTGGAA
401 GAGGATGTGA AATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
451 TGTGCCCCAG TTGGCGGCG GTGACCTGTC AGACATCATC CATGACTTCC
501 AGCGGGGTCT GACTGCGTAC CATGATATCT CCTGGACAA GTGCTATGTC
551 ATCGAACTCA ACACCACCAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
601 CCTCATGAAC GTGAAGAGGG GGACCTACCT GCCGCAGACG TACATCATCC
651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAGGA GGCCCTGGGG
701 TCCTTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGGCCAAG AACTGCAATG
801 CCATCCGCCA CTTCGAGAAC ACCTTCGTCG TGGAGACGCT CATCTCGCGG
851 GTGGTGTGAG GCCCTCCTCC CCCAGAACCC CTTGCCGTGT TCCTCTTTTC
901 TTTCTTCCAG CTGCTCTCTG GCCCTCTCC TTCCCTCTGC TTAGCTTTGA
951 CTTTGGACGC GTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CTGCCCCACC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 TCTGCTGAC CTGGGTGTGG CCGAGGGAGA GGCGATGCTG CAAAGTGTTC
1101 TCTGTGTCCC ACTGTCTTGA AGCTGGGCCT GCCAAAGCCT GGGCCACAG
1151 CTGACCGGCG AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGGCATGTG
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GCGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCTCT AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCTTAGA CCTGGGGCTT GAGCTGAGGA
1351 TGGGACGATG TGCCAGGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCAGGG ACTCTGTAG TGCTTCAGC CCACCAGCAG
1451 GACCTTGGAG TTTGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAAGAGC AAGGAGCTAG GACCCCACT CCTGCCCCC
1551 AGGAGCACAA GCAGGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGG AGGCAAGGTC ACTGCTCAGT CACGTCCACG GGGGACGAGC
1651 CGTGGGTTCT GCTGAGTAGG TGGAGCTCAT TGCTTTCTCC AAGCTTGGAA
1701 CTGTTTTGAA AGATAACACA GAGGGAAAGG GAGAGCCACC TGCTACTTGT
1751 CCACCCCTGC TCCTCTGTTC TGAAATTCCA TCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTCTCT TCTACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGCTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAAACTC TTAATGCTT TGTATATTTT
1951 CTCAATTAGA TCTCTTTTCA GAAGTGTCTA TAGAACATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGGCGG CCG
```

BLAST Results

Entry B64417 from database EMBL:
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.
Length = 715
Plus Strand HSPs:

Medline entries

Peptide information for frame 2

BLASTP hits

Pedant information for DKF2phutel 2h3, frame 2

Report for DKFZphut1_2h3.2

{LENGTH} 267
 {MW} 30253.96
 {pI} 8.16
 {HOMOL} SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
 1e-49
 {PROSITE} MYRISTYL 4
 {PROSITE} PRENYLATION 1
 {PROSITE} CAMP_PHOSPHO_SITE 3
 {PROSITE} CK2_PHOSPHO_SITE 3
 {PROSITE} TYR_PHOSPHO_SITE 1
 {PROSITE} PKC_PHOSPHO_SITE 4
 {PROSITE} ASN_GLYCOSYLATION 1
 {KW} TRANSMEMBRANE 1
 {KW} LOW_COMPLEXITY 15.36 %

SEQ MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY
 SEGxxxxxxxxxxxxxxxx.....
 PRD cccccccchhhhhhhhhhhhhhhhhcccccceccccccccccccccccccchh
 MEMMMM

SEQ LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLYEDSLSSQVRTQMELEEDVKI
 SEG ..xxxxxxxxxxxx.....
 PRD hhhhhhhhhhhhhhhhhhhcchhhhhhhhhccceccccccccchhhhhhhhhhh
 MEM MMMMMMMMMMMMMMMMMMMMM

SEQ YLDENYERINVPVPQFGGDPADIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNF
 SEG
 PRD hhccccccccccccccccchhhhhhhhhhhhhccceccccccccchhhhhhhhh
 MEM

SEQ ELLMNVKRGTYLPQTYIIQEMVTEHVSDEALGSGFIYHLCNGKDTYRLRRRATRRRIN
 SEGxxxxxxxxxxxx.....
 PRD hhhhhccccccccceehhhhhhhccccchhhhhheccccchhhhhhhhhhhhh
 MEM

SEQ KRGAKNCNAIRHFENTFVVETLICGVV
 SEG xx.....
 PRD hhhccccccccchhhhhheeeccc
 MEM

Prosites for DKFZphut1_2h3.2

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	50->54	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PENYLATION	PDOC00266

(No Pfam data available for DKFZphut1_2h3.2)

DKFZphmcfl_1a11

group: transmembrane protein

DKFZphmcfl_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcfl_1a11 encodes a novel 393 amino acid protein, with similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits
potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGGCCAG  CCCCCGAAGA  GCCGCCTCAG  CCGGGGGGAG  TTGCTCGGAC
51  TCAAACTGCC  AGTCCTCGTG  CGACCGCGCT  GGGTCGGAAG  TGAGCAGGCT
101  GAGGCCACCA  TGGAGCAGTG  TCGGTGCGTG  GAGAGAGAGC  TGGACAAGGT
151  CCTGCAGAAG  TTCTGACCT  ACGGGCAGCA  CTGTGAGCGG  AGCCTGGAGG
201  AGCTGTGCA  CTACGTGGGC  CAGCTGCGGG  CTGAGCTGGC  CAGCGCAGCC
251  CTCCAGGGGA  CCCCTCTCTC  AGCCACCCCT  TCTCTGGTGA  TGTACAGATG
301  CTGCCGGAAG  ATCAAAGATA  CGGTGCAGAA  ACTGGCTTCG  GACCATAAAG
351  ACATTACAC  CAGTGTATCC  CGAGTGGGCA  AAGCCATTGA  CAGGAACCTC
401  GACTCTGAGA  TGTGTGGTGT  TGTGTGAGAT  GCGGTGTGGG  ACGGCGGGGA
451  ACAGCAGCAG  CAGATCCTGC  AGATGGCCAT  CGTGGAAAC  CTGTATCAGC
501  AGGCGATGCT  CAGCGTGGCC  GAGGAGCTGT  GCCAGGAATC  AACGCTGAAT
551  GTGGACTTGG  ATTTCAAGCA  GCCTTTCCTA  GAGTTGAATC  GAATCCTGGA
601  AGCCCTGCAC  GAACAAGACC  TGGGTCTCTG  GTTGGAAATG  GCCGTCTCCC
651  ACAGGCAGCG  CCTGCTGGAA  CTCAACAGCT  CCCTGGAGTT  CAAGCTGCAC
701  CGACTGCAC  TCATCCGCCT  CTGGCAGGA  GGCCCCGCGA  AGCAGCTGGA
751  GGCCCTCAG  TATGCTCGGC  ACTTCCAGCC  CTTTGCTCGG  CTGCACCAGC
801  GGGAGATCCA  GGTGATGATG  GGCAGCCTGG  TGTACCTCGG  GCTGGGCTTG
851  GAGAAGTCAC  CCTACTGCCA  CCTGTGGAG  AGCAGCCACT  GGGCAGAGAT
901  CTGTGAGACC  TTTACCCGGG  ACGCCTGTTC  CCTGCTGGGG  CTTTCTGTGG
951  AGTCCCCCT  TAGCGTCAGC  TTTGCCTCTG  GCTGTGTGGC  GCTGCTGTG
1001  TTGATGAACA  TCAAGGCTGT  GATTGAGCAG  CGGCAGTGCA  CTGGGGTCTG
1051  GAATCACAAG  GACGAGTTAC  CGATTGAGAT  TGAATAGGC  ATGAAGTGT
1101  GGTACCACTC  CGTGTTCGCT  TGCCCCATCC  TCCGCCAGCA  GACGTCAGAT
1151  TCCAACCCCT  CCATCAAGCT  CATCTGTGGC  CATGTTATCT  CCCGAGATGC
1201  ACTCAATAAG  CTCATTAATG  GAGGAAAGCT  GAAGTGTCCC  TACTGTCCCA
1251  TGGAGCAGAA  CCGGGCAGAT  GGGAAACGCA  TCATATTCTG  ATTCTACCT
1301  GGAAGGAATT  TTGTTGAAG  GGGTTTTCAC  CTGTGAGCCT  TGGTCTGTCT
1351  CGGTAGGGTG  GTCAACTTCA  GTGGACTGTG  GTTGGTTTCA  GAGCGCCTGG
1401  CTGAGGAGTT  CCACTGAGGG  GAGCACTGGA  GCAGCCCTTT  GGCAGAGGCT
1451  GAGGAGGGAG  ATGGACCAGC  CCACGCCTGG  CACCTGGCTC  CATGGCATAA
1501  GGAAGGGGAG  ATGCTGGCCT  CTGTGCTCCT  GCTGTCTTTT  CCTGTTCTG
1551  TTTGCGTTTG  ACTTAGTAGC  AACCGACAGA  GTGGCAAGGG  ATTTGGTCTT
1601  CAGCAGTAGA  CATCCTTCCA  CCCCTGCCCT  CAGCCAAGTC  TCTTGTGCC
1651  ATGCCAATGC  TATGTCACCC  CTTGCCCTC  GGCCCAAGAG  TGTCCAGCGG
1701  TGGCCACCT  CTTCTCCCA  CTACAGCCTC  AACAGTATGT  ACCATCTCCC
1751  ACTGTAAATA  GTCCCAGTTA  GAACGGAATG  CCGTTGTTTT  ATAACTTTGA
1801  ACAAAATGTA  AAAAAAAAAA

```

BLAST Results

Entry HS579359 from database EMBL:
human STS WI-6350.

Score = 1027, P = 9.9e-40, identities = 207/209

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQCERSLE ELLHYVGQLR AELASAAQOG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDS
101 ICGVVSDAVW DAREQQQIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETFTF DACSLGLSV ESPLSVSFAS GCVALPVLMM
301 IKAVIEQRQC TGVWNHKLDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLCIGHVI SRDALNKLIN GGLKCPYCP MEQNPADGKR IIF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_lal1, frame 2

TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3.
 Length = 398

HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query: 252 YCHLLDSSHWAEICETFTRDACSLGLSVESPLSVSFASGCVALPVLMMIKAVIEQRQCT 311
 Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ + +++++
 Sbjct: 258 YIDVLDLD-WKSLELFLVREFCAALGMSLESPLDIVVNAGAIAPILLKMSSIMKKKHT 316

Query: 312 GVNHHKDELPIEIELGMKCWYHSVFACPI LRQQTSDSNPPIKLCIGHVISRDALNKLING 371
 W + ELP+EI L +HSVF CP+ +Q ++ NPP+ + CGHVI +++L +L
 Sbjct: 317 --WTSQGELPVEIFLPSSYHFSVFTCPVSKEQATEENPPMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKCPYCPMEQNPADGKR IIF 393
 G + KCPYCP E AD R+ F
 Sbjct: 375 GSQRFKCPYCPNENVAADAIRVYF 398

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query: 22 GQCERSLEELLHYVGQLRAELASAAQGTPLSATLSLVMSQCCRKIKD TVQKLASDHD 81
 G C L EL + + + L+ P ++ LV C K + L K
 Sbjct: 15 GNKCLAKLNEL---ESILKDAKKSLKD-PTSMKELVA--CSEKQQVFDLKRTEKK 67

Query: 82 IHSSVSRVGKAIDRNFDS EICGVVSDAVWDAREQQQILQMAIVEHLYQQGMLSVAEELC 141
 H+S++R GK +++ F+ ++ + ++++++ + A+ H ++QG + +A C
 Sbjct: 68 FHTSLNRFGKTLEKKFNFDLEDIKLHSSFESKRE---IDTALS LHF FRQGDVELAHLFC 124

Query: 142 QESTLNVLDLDFKQPFLELN RILEALHEQDLGPALEWAVSHRQRLLELNS SLEFKLHRLHF 201
 +E+ + + F L I++ ++DL +EWA R L SSLE+ L +
 Sbjct: 125 KEAGIEEPSLHVFTLLKSI VQGI RDKDLKPIEWASQCRGYLERKGS SLEYTLQKYRL 184

Query: 202 IRLLAGGPAQL-EALSYAR-HFQPFARLHQREIQVMGSLVY 242
 + K + A+ Y R + F + H +IQ M +L +

Pedant information for DKF2phmcf1_1a11, frame 2

Report for DKFZphmcf1_1a11.2

[illegible]

Prosites for DKF2phmcf1_1a11.2

PS00001	189->193	ASN_GLYCOSYLATION	PDOC00001
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	27->36	TYR_PHOSPHO_SITE	PDOC00007
PS00007	244->253	TYR_PHOSPHO_SITE	PDOC00007
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	50->56	MYRISTYL	PDOC00008
PS00009	387->391	AMIDATION	PDOC00009
PS00013	282->293	PROKAR_LIPOPROTEIN	PDOC00013

545

DKFZphmcf1_1c23

group: mammary carcinoma derived

DKFZphmcf1_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
1 AACTGGCCCC CTCCCCACC CCCTGCCCTC GAGGAGCAGG ACCTGTCCAT
51 GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCTTCAGGCG TCCCAGAGCG TTGTGAGCTC CCGGCTGTCT
151 TCGTCTCTCT CAGCTACTGC TTTGAGATT CAGCCCCCGG GTAGCCAGAG
201 CCTCTCTCCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCCTCCCA GGGAGGACGT AGGTGCTCCC CTGGTCACGC CCTCGCTCTC
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCACCCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
451 GGGCGGGCCA GCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCTGAG GCAGAGCCAC GGCTCCCCA GTCCCTGCC
601 TCAACGGCCA GTTTCATCTT CTCCAAGGGC TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCGTG TCCCCTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGCC CCCAAAGAA
751 TCACCTAAGG CTCCCCACC TGTGGCCCGC AAGCCGCTCT TGGGAGTCCC
801 CCCACCCGCC TCCCCCAGTT ACCCTCGAGC TGAGCCCTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAATGGAG GTGTCTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCCT
951 CCGGGCTCA GACTCACAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGAGCAGCT
1051 CCAAGGACGA GAGGATACAG CAGACACAA CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCT TCGATACTTA TGCAAGCCTG GTGTTGCTCC
1151 TGTCTCTAGA GTCATCTGCG GCTCATGGCT TTTCCCGAAT GGGTTACCT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCTGGA CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTCATCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTGAAGAA GTGGCTGCAT
1401 GGCATCTCTC CAGGGCCAGG GAAGTTGATC TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCTCTCTGC CCAGTCCAGT TTACTCTCTT TGGTCTCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCCACAGGC TTCTTAGGC TGGGGGCAGG TGTGGGGCTG
1601 TGGAAATCCA AAGCACAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCTC
1651 CAACCTACCA ACCACCTCC TGCCCTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCTCTCTT
1801 GACCACACTG GGGTGGGTGG GCCCCAGCA CTTACAGAGG AGGAGCCTTT
1851 GGGGTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCCT TGAGCAGTGG GCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCCTCA TTTGGGTGCA TGTCTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGAAGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TCTAGGGCAG GCTGGGTGGT GGTCCCTTAG GTATCAGCCT
2151 CTCTTACTGT ACTTCCGGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CACCTGTCTA GGCAGCTGG CTTCCTCATT GGCCCTGTG GGTCCACAGC
2251 AGCGTGGCTG CCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTCTCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAGTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCCTAG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC GTTCTGCTC
2601 AGAGCTCTGT ATCTACCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTCCC TCTCCTCCT CCAACATTAC TGGAACTCTA
```

```
2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AARGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AAGTATATAT ATACATATAT ATATTTCTTT
2951 AAATTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAACTGTG TTTCAATTAA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311
Category: putative protein
Classification: unset

```
1 MADFPPEEA FFSVASPEPA GPSGSPELVS SPAASSSSAT ALQIQPPGSP
51 DPPPAPPAPA PASSAPGHVA KLPQKEPVC SKGGGPPRED VGAPLVTPSL
101 LQMVRLRSVG APGGAPTAL GPSAPQKPLR RALSGRASP PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERVSPETQA DLQRNLVAEL RSISEQRPQ APKKSPPKPP PVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGVL QLVGPPEKMG
301 LPGSDSQKEL A
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize
Length = 1,188

HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15
Identities = 81/269 (30%), Positives = 115/269 (42%)

```
Query: 5 PPPEEAFFS----VASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
      PPP S V SP P P SP PA +SS ++ PP +P PPP +
Sbjct: 598 PPPPAPVASPPPPVKSPPPPTPVASPP---PPAPVASSPPPMKSPPPPTPVSSPPPPPEKS 654

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
      PP P PA S P + P P K PP + + P + PS + P
Sbjct: 655 PPPPPPAKSTPPF--EYFT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP 711

Query: 116 PTPALGPSAPQKPLRRA-LSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
      P+ PS P++P+ + ++SP PAP S +LA S + + PP
Sbjct: 712 PSSPEKPSPPKEPVSSPPQTPKSSPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPQAPK 233
      PP +P +S +Q+ P +P++ L V+ + + PP AP
Sbjct: 772 PLSSPPAPQVKSS-----PPPVQVSSPPAPKSSPPLAP--VSSPPQVEKTSPPAPL 823

Query: 234 KSPKAPPPVARKPSVGV--PPPASPSYFRAEPLTAPPTNGLP 273
      SP P + P V V PPP S P P+++PP P
Sbjct: 824 SSPPLAPK-SSPHVVVSSPPVVKSSPPAPVSSPPLTPKP 864

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14
```

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPVLSSPAASS---SSATALQIQPPGSPDPPAP---PAPAPASSAPGHV 69
 P P G P S P + P A A S + S T + P P + P P P P P + P
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSHTPPDVSPPELPEPSVPAPAPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQKP 128
 +P PV G S P V P + +V+L AP G+P P + +P P
 Sbjct: 469 DYVPPTPPVPGKSPATSPSPQVQPPAASPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188
 + G S P P S + +K+ A G + P PPE P PP AS
 Sbjct: 529 I-----GSPSP-PPVSVVSPPPPVKSPPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSKGRKLQLERPVSPTQADLQRLNVAELRSISEQRPPQAPKKSPKAPPPVARKPS- 247
 + S L P P ++ VA + PP P SP P PVA P
 Sbjct: 578 PVKSPPPPTLVASPP--PPVKSPPPPAPVASPPPVKSPPPTPVASPPPPAPVASSPPP 635

Query: 248 VGVPPP---ASPSYPRAEPLTAPPTNGLPHTQD 277
 + PPP +SP P P PP P ++
 Sbjct: 636 MKSPPPTPVSSPPPEKSPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPVLSSPAASS---SSATALQIQPPGSPDPPAPAPAPASSAPGHVA 70
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+
 Sbjct: 817 SPPPA-PLSSPPLAPKSSPPHVVSPPPPVKSPPPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ---KEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQ 126
 P+ P + PP E +P TP L ++S P +P + P +
 Sbjct: 873 SPPEVVKPSTPPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPAMVSSPMTPKSSP 932

Query: 127 KPLRRAL---SGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P
 Sbjct: 933 PPVVSSPPPTVKSSPPAPVSSPPATP--KSSPPAPVNL---P--PPEVKSPPPTP 984

Query: 184 ASTASFIFSKGRKLQLERPVSPTQADLQRLNVAELRSISEQRPPQAPKKSPKAPPPVA 243
 S+ + P PE ++ V+ + PP AP SP PPV
 Sbjct: 985 VSSPPAPKSSPPAPMSSPPPEVKSPPPPAPVSSPPPVKSPPPPAPVSSP--PPPVK 1042

Query: 244 RKPS---GVPPPPASPSYPRAEPLTAPP 268
 P V PPP S P P++PP
 Sbjct: 1043 SPPPPAPVSSPPPVKSPPPPAPISPP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSPAAS---SSATALQIQPPGSPDPPAPAPAPAS 63
 PPP S PE + P P + P + T+++ PP PP P+P
 Sbjct: 639 PPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPPPTLIPSPPP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSP 123
 P K P K PP+E V +P TP V +P PTP P
 Sbjct: 699 QEKTTPSTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPAPVSSP--PPTPVSSPP 753

Query: 124 APQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
 A P+ S ++SP PAP S A ++K+ + + P PP + PP +P
 Sbjct: 754 A-LAPVSSPPSVKSSPPAPLSSPPAPQVKS----SPPPVQVSSP--PPAPKSSPPLAP 806

Query: 184 ASTASFIFSKGRKLQLERP-VSPETQADLQRLNVAELRSISEQRPPQAPKKSPKAPPPV 242
 S+ + L P ++P++ +V+ + + PP AP SP P
 Sbjct: 807 VSSPPQVEKTSPPAPLSSPPLAPKSSPP--HVVVSSPPPVKSSPPPPAPVSSPPLTPK 864

Query: 243 ARKPS-VGVPP---PASPSYPR-----AEPLTAPP 268
 A P+ V PP P++P P +EP ++PP
 Sbjct: 865 ASPPAHVSSPPEVVKPSTPPAPTTVISPPSEKSSPP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSPAAS---SSATALQIQPPG---SPDPPAP--- 56
 PPP A S P P S P + VSSP A SS A PP PPPAP
 Sbjct: 768 PPP--APLSSPPAPQVKSSPPPVQVSSPPAPKSSPPLAPVSSPPQVEKTSPPAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAP 116
 P AP SS P V P PV S PP V +P +TP V +P
 Sbjct: 826 PPLAPKSSPPHVVSPP--PVVKSS---PPAPVSSPPLTPKAPSPA--HVSSPPEVV 878

Query: 117 TPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQ--- 169
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPAPTTVISPPSEKSSPPPTPVSLPPPVIKSSPPAMVSSPMTPKSSPPPVV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIKSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRP 228
+ PP + PP +P S+ + P PE ++ V+ + P

Sbjct: 938 SSSPPTVKSSPPAPVSSPPATPKSSPPAPVNL P-PPEVKSSPPPTPVSSPPAPKSSP 996

Query: 229 PQAPKKSAPKPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPPEVKSSPPAPVSSPPPVKSSPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGPSGSELVSSPAASSSATALQIQPP---GSPDPP---PA 55
D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKL----PQKEPVGCSKGGGPPREDVGAPLVTPLQLQMVRLRSVGA 111
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPPVSVSSPPPVKSSPPAPVGSPPPEKSPPPAPVASPPPVKSSPP 584

Query: 112 PG--GAPTPALGPSAPQKPLRRA---LSGRASVPAPSSGLHAAVRLKACSLAASEGLSS 166
P +P P + P P+ + P P S A V + + +

Sbjct: 585 PTLVASPPPVKSSPPAPVASPPPVKSSPPPTPVASPPAPVASSPPPMKSPPPPTP 644

Query: 167 AQPNGPPEAEPRPPQSPASTASFIKSGSRKLQLERPVSPETQADLQRNLVAELRSISEQ 226
PPE P PP PA + + ++ PE L+ +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYTPPTS VKSSPPPEKSLP-PPTLIPSPPPQEK 702

Query: 227 RPPQAPKKSAPKPP-PVARKPSVGVPPASPSYPRAEPLTAPP 268
PP P K P +P P K V PP S P P+++PP

Sbjct: 703 TTPSTPSKPPSSPEKSPPEKSPVSSPPQTPKSSPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAEFFSVASPEPAGP---SGSELVSSPAASSSATALQIQPPGSP--DPPAP-- 56
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPAPVSSPPLTPKAPASPPAHVSSPPEVVK-PSTPPAPTTP--ISPPSEKSSPPPTPV 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLQLQMVRLRSVGA PGGA 115
P P SS P + P P P PP V +P P++ V +P

Sbjct: 907 LPPPIVKSSPPAMVSSPMTPKS-----SPPVVVSSP--PPTVKSSPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 PKSSPPAPVNLPPPEV---KSSPPPTPVSSPPAPK----SSPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIKSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPQAPKKS 235
+ PP +P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPAPVSSPPPVKSSPPAPVSSP-PPPVKSSPPAPVSSPPPVKSSPPAPISS 1068

Query: 236 PKAPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P--PPPVKSSPPAPVSSPPPVKSSPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPPGSPDPPAP---PAPAPASSAPGHV 69
P P G P SP + PAAS+ S T + P P+P P P P P +P

Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSHTPPDVSPEPLPEPSVPAPAPMMPMTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLQLQMVRLRSVGA PGGAPTALGPSAPQK 128
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVSSPPPVKSSPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSGSRKLQLERPV---SPETQADLQRNLVAELRS-----ISEQRPQA-----PK 233
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSPPPPTLVASPPPVKSSPPAPVA-SPPPVKSSPPPTPVASPPAPVASSPPPM 636

Query: 234 KSPKAPPPVARKP---SVGVPPASPSPYPRAEPLTAPPTN 270
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPTPVSSPPPEKSSPPPPPAKSTPPPEEYTPPTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFSVASPEPAGSGSPSELVSSPAASSSSATALQIQQSPDPPAPPAPAPASS 64
 PP S S + P + P + P SS A+ PP +P +PP P SS
 Sbjct: 883 PPAPTVISPPSEPKSSPPPTPVSLPPPIVKSSPPAMVSSPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPG--GAPTALGP 122
 P V P PV PP +P P L ++S P +P PA
 Sbjct: 940 PPFTVKSSPPAPVS-----SPPATFKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPP 180
 S P P+ ++ P PAP S V+ S +SS P PP + PP
 Sbjct: 995 SPPAPMSSPPPEVKSPPPAPVSSPPPVK----SPPPPAPVSS--P--PPPVKSPP 1046

Query: 181 QSPASTASFIKSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKSPKAPP 240
 +P S+ + P P ++ V+ + PP AP SP PP
 Sbjct: 1047 PAPVSSPPPVKSPPPPAPISSP-PPPVKSPPPPAPVSSPPPVKSPPPPAPVSSP--PP 1103

Query: 241 PVARKPS---VGVPPPAS---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283
 P+ P V PPPA TS P P+++PP P + ++ L
 Sbjct: 1104 PFKSPPPPAPVSSPPAPVKPSLPPAPVSSPPPVTPAPPKKEEQL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFSVASPEPAGSGSPSELVSSPAASSSSATALQIQQP----GSPDPP---PA 55
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +
 Sbjct: 469 DVPPTTPP---VPGKSPFATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115
 PPAP + S P V+ + PV PP VG+P P V+P
 Sbjct: 525 PPAPIGSPSPPPVSVVSPPPPVKSP----PPAPVGSPP--PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175
 P P P P ++ P PAP + V+ S ++S P P +
 Sbjct: 576 PPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPVK----SPPPTPVASPPPPAPVAS 631

Query: 176 EPRPPQSPASTASFIKSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
 P P +SP K P P S+ PP+
 Sbjct: 632 SPPPMKSPPPPTPVSSPPPEKSP--PPPPAKSTPPPEYPTPTSVKSSPPPEKSLPP 689

Query: 236 PK---APPPVARK--PSVGVPPPASPSYPRA--EPLTAPP 268
 P +PPP + PS PP+SP P EP+++PP
 Sbjct: 690 PTLIPSPPPQEKPTPSTPSKPPSSPEKPSPPKEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPPEEAFSVASPE-PAGPSGSPSELVSSPAASSSSATALQIQQSPDPP-PAPPAP 59
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P
 Sbjct: 496 ASTPPP--SLVKLSPPQAPVGSPPPPVKTTSPPAPIGSPSPPPVSVVSPPPPVKSPPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGAPTPA 119
 AP S P P PV PP + P + S V+ AP +P P
 Sbjct: 554 APVGSPPPEKSPPPAPVASPP--PPVKSPPPTLVASPPPPVKSPPPPAPVASPPPP 610

Query: 120 LGFSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSL-AAASEGLSSAQNGPPEAEPR 178
 + P P+ + P PAP + ++ +S P PP A+
 Sbjct: 611 VKSPPTTPVA-----SPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKST 664

Query: 179 PP--QSPASTASFIKSGSRKLQLERPV--SPETQADLQRNLVAELRSISEQRPPQAPK 233
 PP + P S S K L P SP Q S ++P +P
 Sbjct: 665 PPPEEYPTPTTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPSTPSKPPSSPEK--SPP 721

Query: 234 KSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 K P + PP K S PPPA S P P+++PP
 Sbjct: 722 KEPVSSPPQTKSS---PPAPVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGSGSPE-LVSSPAASSSS---ATALQIQQSPDPP-- 54
 PPPE++ VASP P S P LV+SP S A PP PPP
 Sbjct: 560 PPPEKSPPPAPVASPPPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPTP 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVGC---SKGGGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP +S+P + P PV K PP P ++S
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKSTPPPEEYPTPTSVKS 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSA 167
 P + P P L PS P P + + ++P PSS + + S SS
 Sbjct: 680 SPPPEKSLPPTLIPSP--PQEKPTPSTPSKPPSSPEKPSPPKEPVSSPPQTKSSP 736

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSGSRKLQLERPVS PETQADLQRNLVAELRSISEQR 227
P P P SP + A + S S K P + P + + + +
Sbjct: 737 PPAPVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPVQVSS 793

Query: 228 PPQAPKKSPPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
PP APK SP P+A P V PP + P PL++PP
Sbjct: 794 PPPAPKSSP----PLA--P-VSSPPQVEKTSPPPAFLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09
Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQQPGSPDPPPP-APPAPA 60
PPP + + + P P G PS P +VS P S P GSP PP +PP PA
Sbjct: 517 PPPVK---TTSPPAPIGSPSPPPPVSVSPPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP--LVTPSLLQMVRLRSVGAPGG 114
P +S P V P P V PP V +P + +P V AP
Sbjct: 571 PVASPPPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASVPVAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171
+ P + P P+ SP P P S+ S+ +S + P
Sbjct: 631 SSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQA 231
PP P PP T SK P SPE + + V+ + PP A
Sbjct: 689 PPTLIPSPPPQEKPTPSTPSKP-----PSSPEKPS-PKEPVSSPPQTPKSSPPPA 739

Query: 232 PKKSPPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268
P SP P PV+ P++ PP+ S P PL++PP
Sbjct: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPPAFLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08
Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQQPGSPDPPPPAPPAPA 60
A P P SPEP PS P P + S A PP P P +PPA +
Sbjct: 427 ASAPMPSPHTPPDVSPPELPEPSVPVAPAPMPMPTPHSPPADDDYVPPTPPVPGKSPFATS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT-- 118
P+ A P V S PP+ VG+P P V+ S AP G+P+P
Sbjct: 487 PSPQVQPPAASTPPPSLVKLS----PPQAPVGSPP--PPP---VKTSPAPAGISPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174
+ P P K P A G SP P S A S + + PP
Sbjct: 537 PVSVPSPPPVKSPPPPAPVG--SPPPPEKSPPPPAPVASPPPPVKSPPPTLVASPPPP 594

Query: 175 AEPRPPQSPASTASFIFSGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKK 234
+ PP +P ++ + P P A + + PP P+K
Sbjct: 595 VKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPPTNGLP 273
SP PPP P PP P+ P + + PP LP
Sbjct: 654 SPPPPPPAKSTP---PPEEYPTPTSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08
Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQQPGSP--DPPAP--PAP 59
PPP V+SP P P SP P SS ++ PP +P PP P P P
Sbjct: 916 PPPA---MVSSP-PMTKSSPP---PVVVSSPPPTVKSSPPAPVSSPPATPKSSPPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTAP 119
AP + P V P PV S P AP+ +P + V+ AP +P P
Sbjct: 967 APVNLPPPEVKSSPPPTVVS-SPPPAKSSPPAPMSSPPPE-VKSPPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPVAPSSGLHAAVRLKACSLAASEG---LSSAQNGPPEA 175
+ P P+ ++ P PAP S V+ S + S P P +
Sbjct: 1025 VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKKS 235
P P +SP A S ++ P P A + A ++ S PP AP S
Sbjct: 1085 PPPPVKSPPPPAPV---SSPPPIKSPPPP---APVSSPPAPVKKPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYRAEPLTAPP 268
P P +K +PPA S P + PP
Sbjct: 1136 PPPVTPAPPKKEQSLPPPAESQPPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06
Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFPPEEAFSSVASPEP-AGPSGSELVSSPAASSSSATA-LQIOPPGSP--DPPP--A 55
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +
 Sbjet: 970 NLPPPEVK--SSPPPTVSSPPAPKSSPPAPMSSPPPEVKSSPPPPAPVSSPPPPVKS 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115
 PP PAP SS P V P PV PP + P S V+ AP +
 Sbjet: 1028 PPPPAPVSSPPPPVKSPPPPAPVSSPP--PPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 116 PTPALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174
 P P + P P+ ++ P PAP S A +K SL +SS P PP
 Sbjet: 1085 PPPPVKSPPPPAPVSSPPPIKSPPPAPVSSPPAP-VKPPSLPPPPAPVSS--P--PPV 1139

Query: 175 AEP RPPQ 181
 P PP+
 Sbjet: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPEEAFSSVASPEPAGP-SGSELVSSP--AASSSSATALQIOPPGSP--DPPP 54
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP
 Sbjet: 1001 MSSPPPE-----VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPP 1055

Query: 55 ---APPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP SS P V P PV PP V +P P +
 Sbjet: 1056 PVKSPPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125
 V +P AP P+L P AP
 Sbjet: 1114 VSSPPPPAPVKPPSLPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFSS---VASPEPAGP-SGSELVSSP--AASSSSATALQIOPPGSP--DPPP 54
 PPP S V SP P P S P V SP A SS ++ PP +P PPP
 Sbjet: 1060 PPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPIKSPPPAPVSSPPPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108
 AP P PAP SS P V P K+ + PP E P +L +
 Sbjet: 1120 APVKPPSLPPPPAPVSSPPPPVTPAPPKKE---EQSLPPPAESQPPPSFNDIILPIMANK 1176

Query: 109 VGAP 112
 +P
 Sbjet: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQNG 171
 G PTP GP + P + A S +P+P+ + L S + A + P+
 Sbjet: 408 GYPTGGGGPPSSPVGKPAAS---APMPSPHTPPDVSEPLEPEPSVPVAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFIFSKGRKLQLERPVSPEQ---ADLQRNLVAELRSISEQR 227
 PP + PP P S + S ++Q +P + Q + + +
 Sbjet: 465 PPADDYVPPTPPVPGKSPATSPSPQVPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 228 PPQAPKKSPPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 PP AP SP PPPV SV PPP S P P+ +PP
 Sbjet: 525 PP-APIGSPSPPPV---SVVSPPPVKSPPPPAPVGSPP 560

Pedant information for DKFZphmcf1_lc23, frame 1

Report for DKFZphmcf1_lc23.1

[LENGTH] 311
 [MW] 31534.58
 [pI] 9.48
 [KW] All Alpha
 [KW] LOW_COMPLEXITY 38.59 %

SEQ MADFPPEEAFSSVASPEPAGPSGSELVSSPAASSSSATALQIOPPGSPDPPPPAPPAPA
 SEGXX
 PRD ccc
 SEQ PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTAL
 SEG xxxxxxxx.....xx

```
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      GPSAPQKPLRRALSGRASVPVPAPSSGLHAAVRLKACSLAASEGLSSAOPNGPPEAEPRPP
SEG      xxxxx.....xxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ      QSPASTASFIFSKGSRKLQLERPVPETQADLQRNLVAELRSISEQRPPQAPKKSPPKAPP
SEG      xxxxx.....xxxxxxxxxxxxxx
PRD      cccccceeeccccChhhhccccccccchhhhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ      PVARKPSVGVPPPASPSYPRAEPLTAPPTNGLPHTQDRTKRELAENGVLQLVGPEEKMG
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhcccccecccccccccc

SEQ      LPGSDSQKELA
SEG      .....
PRD      ccccccccccc
```

(No Prosite data available for DKFZphmcf1_1c23.1)

(No Pfam data available for DKFZphmcf1_1c23.1)

DKF2phmcf1_le15

group: transmembrane protein

DKF2phmcf1_le15 encodes a novel 454 amino acid protein with similarity to C. elegans proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER
membrane regions: 9

complete cDNA, complete cds, EST hits
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp

Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGCAGCGC  CCGGGCTGAG  CGACAGCAAG  TGCAGCGGGC  TCCTACCCCG
51  GGTGAGGGGT  GGCCTCCGCG  TGGGATCGTG  CCTCTTTCAG  CCCGCTCCTG
101  TCCCCGACAT  CAGCTGTATT  CCGCAGCTCC  CCTCCGCGCT  GTGTGCTTAC
151  TGAGACGGGG  AGCGGTGACA  GGGCCCGGGT  CCTTCTCAG  TGGTGTCTG
201  TGCTTCAGGG  CAAGCTCCCC  GTCTCCGGGC  GCACTTCCCT  CGCTGTGTT
251  CGGTCCATCC  TCCTTCTTCC  AGCCTCCTCC  CCTCGCAGGT  GGGATCGTCG
301  GTGGGACCGG  AGCGCGGGCG  GCGCGGGCCC  CCCGGGACCA  TGGCCGGGTC
351  CGACACCGCG  CCTTCTCTCA  GCCAGGCGGA  TGACCGGGAC  GACGGGCGAG
401  TGCTTGGCAC  CCGGGGTTG  CCAGGGTCCA  CGGGGAACCC  GAAGTCCGAG
451  GAGCCGAGG  TCCCGGACCA  GGAGGGGCTG  CAGCGCATCA  CCGGCTGTG
501  TCCCGGCGGT  TCGGCTCTCA  TAGTGGCGGT  GCTGTGTACT  ATCAATCTCC
551  TGAATACAT  GGACCGCTTC  ACCGTGGCTG  TGTTTATCTC  CAGTTACATG
601  GTGTTGGCAC  CTGTGTTTGG  CTACCTGGGT  GACAGGTACA  ATCGGAAGTA
651  TCTCATGTGC  GGGGGCATTG  CCTTCTGGTC  CCTGCTGACA  CTGGGGTCAT
701  CCTTCATCCC  CGGAGAGCAT  TTCTGGCTGC  TCCTCCTGAC  CCGGGGCTG
751  GTGGGGTTCG  GGGAGGCCAG  TTATTCCACC  ATCGCGCCCA  CTCTCATTGC
801  CGACCTCTTT  GTGGCCGACC  AGCGGAGCCG  GATGCTCAGC  ATCTTCTACT
851  TTGCCATTCC  GGTGGGCACT  GGTCTGGGCT  ACATTGCAGG  CTCCAAAGTG
901  AAGGATATGG  CTGGAGACTG  GCACTGGGCT  CTGAGGGTGA  CACCGGCTCT
951  AGGAGTGGTG  GCCGTCTGCG  TGCTGTTTCT  GGTAGTGCAG  GAGCCGCCAA
1001  GGGGAGCCGT  GGAGCGCCAC  TCAGATTTCG  CACCCCTGAA  CCCCACCTCG
1051  TGGTGGGCAG  ATCTGAGGGC  TCTGGCAAGA  AATCTCATCT  TTGACTCAT
1101  CACCTGCCTG  ACCGGAGTCC  TGGGTGTGGG  CCTGGGTGTG  GAGATCAGCC
1151  GCGGGCTCCG  CCACTCCAAC  CCGCGGGCTG  ATCCCTGGT  CTGTGCCACT
1201  GGCTCCTGCG  GCTCTGCACC  CTCTCTTCT  CTGTCCCTTG  CTGTGGCCCG
1251  TGGTAGCATC  GTGGCCACTT  ATATTTTCAT  CTTTATTGGA  GAGACCTTCC
1301  TGTCATGAA  CTGGGCCATC  GTGGCCGACA  TTCTGTGTGA  CTGTGTGATC
1351  CCTACCCGAC  GCTCCACCGC  CGAGGCCCTT  CAGATCGTGC  TGTCCACCT
1401  GCTGGGTGAT  GCTGGGAGCC  CCTACCTCAT  TGGCTGTATC  TGTACCCGCC
1451  TGCGCCGGAA  CTGGCCCCCC  TCCTTCTTGT  CCGAGTTCCG  GGCTCTGCAG
1501  TTCTCGCTCA  TGCTCTGCGC  GTTGTGTGGG  GCACTGGGCG  GCGAGCCCTT
1551  CCTGGGCACC  GCCATCTTCA  TTGAGGCCGA  CCGCGGGCGG  GCACAGCTGC
1601  ACGTGCAGGG  CTGTCTGCAC  GAAGCAGGGT  CCACAGACGA  CCGATTGTG
1651  GTGCCCCAGC  GGGGCCGCTC  CACCCGCGTG  CCCGTGGCCA  GTGTGCTCAT
1701  CTGAGAGGCT  GCCGCTCACC  TACCTGCACA  TCTGCCACAG  CTGGCCCTGG
1751  GCGCACCCCA  CGAAGGGCCT  GGGCCTAACC  CCTTGGCCTG  GCCCAGCTTC
1801  CAGAGGGACC  CTGGGCGGTG  TGCCAGCTCC  CAGACACTAC  ATGGGTAGCT
1851  CAGGGGAGGA  GGTGGGGGTC  CAGGAGGGGG  ATCCCTCTCC  ACAGGGGCAG
1901  CCGCAAGGGC  TCGGTGCTAT  TTGTAACGGA  ATAAAATTTC  TAGCCAGAAA
1951  AAAAAAA

```

BLAST Results

Entry E12646 from database EMBL:
cDNA encoding cell growth inhibiting factor.
Score = 3046, P = 2.2e-131, identities = 640/659

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454
 Category: similarity to known protein

```

1 MAGSDTAPFL SQADDPDDGP VPGTPGLPGS TGNPKSEEP VPDQEGLRQI
51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISYMYLA PVFGYLGDRY
101 NRKYL MCGGI AFWSLVTLGS SFIPGEHFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRMLSIYF FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LFLVVRPPEP RGAVERHSDI PPLNPTSWWA DLRALARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRD DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCAFGALG
401 GAAFLGTAIF IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1e15, frame 1

TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,
 N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid
 C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,
 N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11_18 gene: "F6H11.180"; product: "predicted protein";
 Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII
 project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT_LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N
 = 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9
 Length = 488

HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISSYMYLAPVFGYLGDRYNRKYL MCGGIAFWSLVT 117
 + ++ V Y N + + + VF + S + MV + PV GYLGD + NRK + + M G + W
 Sbjct: 29 AGVLTVQVQTYYNISDSLGGLIQTVFLISFMVFSVPCGYLGDRFNRRKIMIIIGVIGLAV 88

Query: 118 LGSSFIPGEHFWLLLLTRGLVGVGEASYSTIAPTLLIADLFVADQRSRMLSIYFIFAIPVGS 177
 LGSSF + P HFWL L + R VG + GEASYS + AP + LI + D + F + RS + IFYFAIPVGS
 Sbjct: 89 LGSSFVPANHFVLFVLRSFVIGGEASYSNVAPSLISDMFNGQKRSTVFIMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRPPEPRGAVER --- HSDLPPL 233
 GLG + I GS V + G W W + RV + G + + L L EP RGA + + D +
 Sbjct: 149 GLGFIVGSNVATLTGHMQGWIRVSATAGLIVMIALVLFTEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWWADLRALARNLIFGLITCLTG 259
 T + + DL L + L + C G
 Sbjct: 209 TNTTYLEDLVILLKTPT -- LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGLVGLGVEISRRL ----- RHSNPRADPLVCATGLLGSAPFLFLSL 300
 L FG IT G + + GV G + S + L R RA PLV G L + APFL + +
 Sbjct: 277 LYFGAITTAGGLIGVIFGSMLSKWLAVAGWGPFRRLQTDRAQPLVAGGALLAAPFLIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

```

SV YI IF G T + NW + D+L V + P RRSTA ++ +++SHL GDA
Sbjct: 337 IFGDKSLVLLYIMIFGFIGTFMCFNWGLNIDMLTTVIHPNRRSTAFSYFVLVSHLFGDASG 396

Query: 361 PYLIGLISDLRLRN--WPPSFLSEFRALQFSMLCAFVGALGGAFLGTAFIEDRR-- 416
      PYLIGLISD +R + + +L + C + L + + + + + +DR+
Sbjct: 397 PYLIGLISDAIRHGSTYPKD--QYHSLVSATYCCVALLLSAGLYFVSSLTLVSDRKKF 453

Query: 417 RAQLHVQGLLHEA--GSTD--DRVVPQGRGRSTRV 447
      RA++ + L + STD +RI + S+R+
Sbjct: 454 RAEMGLDDLQSKPIRSTSTDSLERIGINDDVASSRL 488

Score = 70 (10.5 bits), Expect = 5.9e-24, Sum P(2) = 5.9e-24
Identities = 25/89 (28%), Positives = 41/89 (46%)

Query: 62 VAVLCYINLLNYMDRFTVAVFISSYMVLPAPVFGYLGDRYNRKYLMCGGIAFWSLVT--LG 119
      V L +NLLNY+DR+TVA ++ + LG +L+ +S V LG
Sbjct: 11 VTLFVLLNLYVDRTVAVAGVLTVQVQTYNIDSLSGLGIQTVFLLI--SFMVFPVCGYLG 68

Query: 120 SSFIPGEHFVLLLLTRGLVGVGEASYSTIAP 150
      F W++++ G +G S S P
Sbjct: 69 DRF--NRKWIMIGVG-IWLGAVLGSSFP 95

```

Pedant information for DKFZphmcfl_1e15, frame 1

Report for DKFZphmcf1_1e15.1

```
[LENGTH]          454
[MW]               49013.35
[pI]               7.66
[HOMOL]            TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-S1

[BLOCKS]           BL01022D
[PROSITE]          MYRISTYL             11
[PROSITE]          CAMP_PHOSPHO_SITE     1
[PROSITE]          CK2_PHOSPHO_SITE      3
[PROSITE]          PROKAR_LIPOPROTEIN    1
[PROSITE]          GLYCOSAMINOGLYCAN    1
[PROSITE]          PKC_PHOSPHO_SITE      4
[KW]               TRANSMEMBRANE       8
[KW]               LOW COMPLEXITY        15.42 %
```

[illegible]

WO 01/12659

PCT/IB00/01496

SEG
PRD hhhhhhhhhcccccccccccccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

Prosite for DKFZphmcfl_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcfl_1e15.1)

DKF2phmcf1_lg13

group: mammary carcinoma derived

DKF2phmcf1_lg13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKF2

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```

1  GAAACCTGAT CTCATAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51  GACCCATATT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACCTACA GATTGTCTC CAGGGGGTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
251 CTACAACTAT GAATGAGAGA CCCTTATTGT CATCGTATT AGTTGCATAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA GCGGCTGAAA AAATATCCT
351 TCCAGCATGT ATGGACATGG TAGGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATTG GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCACACTC CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGT
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAAC
701 GGAAACATTG TAAAGGAATT TCAAGTGATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACATGC
801 TGTTTGGAAT CACTGTTTTA TTCATCGAGA AGCTTTGGTA TCCAAAGAAA
851 TTTACCAAG TCTGATGGAT GTATTGAAA ATGCAAGTGA AACTGTTAAT
901 TTTATTAAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTGTTC
951 AGAGATTGGA GTGAACCACA CCCACTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAC CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTGGCAA ATATTTTGA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTAT ATTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAAAAACAA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC CCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGCTCAAAA CTTTAAATTA
1401 TTACTTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAAAT CAATAATTGA GTTAAACTTG
1501 GAGCCTGAAG AAGAGAAATGA ATTATTGCAG CTCAGTTCAT CATTACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACCTA CATATTGTG TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAACAA AAGAGAAATA GGCTCAATAG TGACACAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTTCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCCATCAC ATTAATAACA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAC CCCATCTCTA
1951 CTAAAAAATAG AACCTTAGC CAGGCGTGGT GGCACATGCC TGCAAGTCCA
2001 GTTACTTGGG TGCTGAGGC AGGAGAACTC CTTAAACAG GAAGGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGTCCA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAATTGT ATTTGACTTT TTAAGGGAT
2151 TTTGCAGTAT GTTGTAGTTA AACGTTAATA AAATTATATT TGTAATTAGG
2201 AAAAAAAAAA
```

BLAST Results

Entry AC005020 from database EMBL:
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.
Score = 9110, P = 0.0e+00, identities = 1822/1822

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573
 Category: similarity to unknown protein

```

1 MTPESRDITD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
51 SRSTTMNERA LLSSYLVAIR VAKKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKHKCGIS SDGTANMTGK HSRLTEKLE ATHNNVWNN CFHREALVS
251 KEISPSLMDV LKNAVKTVMF IKGSSLNSRL LEIFCSEIGV NHTHLLFHTF
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFQK TLLMQARLK SNRPSYMF
401 TLLQHIEENI INEDCLKEIK LEILLHLTSL SQTFNYYFPE EKFECLKENI
451 WMKDPFAFQN PESITELNLE PEEENELLQL SSSFTLNYY KILSLAFWI
501 KIKDDFPLLS RKSILLLLPF TTYLCELGF SILTRLTKK RNRLNSAPDM
551 RVALSSCVPD WKELMNRQAH PSH

```

BLASTP hits

Entry AC004877_3 from database TREMBLNEW:
 gene: "WUGSC:H_DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens
 PAC clone DJ0751H13 from 7q35-qter, complete sequence.
 Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211_1 from database TREMBL:
 product: "Hermes transposase"; Musca domestica Hermes transposase
 gene, complete cds.
 Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

Alert BLASTP hits for DKFZphmcf1_lg13, frame 1

TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P = 1.1e-23

>TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds.
 Length = 607

HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23
 Identities = 120/485 (24%), Positives = 229/485 (47%)

```

Query:      89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
             CM+ ++R + + L + LS + +RI +I ++L L R + +++ LD+
Sbjct:     124 CMEVLLREVLPFH-VSVLQGVOLSPDITRQILSIDRNLRLNQLFNRRDFKAYSLALDDQ 182

Query:     148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
             +A LLV++R V + + EDLL +NL H + G + LE+ L L + +
Sbjct:     183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHFVSGALMSAILES--LQTAGLSLQR 240

Query:     206 CKGISSDGTANMTGKHSRLTEKLEATHNNVWNN--HC--FIHREALVSKEISPSLMDVL 261
             G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct:     241 MVGLTTTHTLRMIGENSGLSVMREKAVSPNCWNVIHYSGLHLELLSSYQVDVN--QII 298

Query:     262 KNAVKTVMFIKGSLSLSRLLEIFCSEIGVNNHTHLLFHTFVR-WLSQGVLSRVYELRNEI 320
             + + IK + + +E H + + WL +GK L ++ LR E+
Sbjct:     299 NTISEMIVLIKTRGVRRPEFQTLTTESESEHGERVNGRCNLNWLRRGKTLKLIFSLRKEM 358

Query:     321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFGILNELSLKMQGKNNDIFQYLEHILGFQK 380
             FLV + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:     359 EAFLVSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417

```

Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22
Identities = 120/485 (24%), Positives = 228/485 (47%)

Pedant information for DKFZphmcf1_lg13, frame 1

Report for DKFZphmcf1 1q13.1

```

[LENGTH]          573
[MW]               66276.85
[pI]               5.82
[HOMOL]            TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens
mRNA for KIAA0766 protein, complete cds. 1e-18
[PROSITE]          MYRISTYL             3
[PROSITE]          CK2_PHOSPHO_SITE      10
[PROSITE]          TYR_PHOSPHO_SITE      1
[PROSITE]          PKC_PHOSPHO_SITE      9
[PROSITE]          ASN_GLYCOSYLATION     2
[KW]               All_Alpha
[KW]               LOW_COMPLEXITY        8.90 %

```

```

SEQ      MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDDEEDHFQKERNKVESSPQVLSRSTTMNERA
SEG      .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      LLSSYLVAYRVAKEKMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRC

```

[illegible]

Prosites for DKFZphmcf1_1g13.1

PS000001	216->220	ASN GLYCOSYLATION	PD00000001
PS000001	291->295	ASN GLYCOSYLATION	PD00000001
PS000005	116->119	PKC PHOSPHO_SITE	PD00000005
PS000005	218->221	PKC PHOSPHO_SITE	PD00000005
PS000005	225->228	PKC PHOSPHO_SITE	PD00000005
PS000005	358->361	PKC PHOSPHO_SITE	PD00000005
PS000005	391->394	PKC PHOSPHO_SITE	PD00000005
PS000005	445->448	PKC PHOSPHO_SITE	PD00000005
PS000005	485->488	PKC PHOSPHO_SITE	PD00000005
PS000005	510->513	PKC PHOSPHO_SITE	PD00000005
PS000005	538->541	PKC PHOSPHO_SITE	PD00000005
PS000006	55->59	CK2 PHOSPHO_SITE	PD00000006
PS000006	79->83	CK2 PHOSPHO_SITE	PD00000006
PS000006	95->99	CK2 PHOSPHO_SITE	PD00000006
PS000006	136->140	CK2 PHOSPHO_SITE	PD00000006
PS000006	183->187	CK2 PHOSPHO_SITE	PD00000006
PS000006	189->193	CK2 PHOSPHO_SITE	PD00000006
PS000006	256->260	CK2 PHOSPHO_SITE	PD00000006
PS000006	445->449	CK2 PHOSPHO_SITE	PD00000006
PS000006	463->467	CK2 PHOSPHO_SITE	PD00000006
PS000006	546->550	CK2 PHOSPHO_SITE	PD00000006
PS000007	364->372	TYR PHOSPHO_SITE	PD00000007
PS000008	137->143	MYRISTYL	PD00000008
PS000008	273->279	MYRISTYL	PD00000008
PS000008	289->295	MYRISTYL	PD00000008

(No Pfam data available for DKFZphmcf1_lgl3.1)

DKFZphtes3_14g5

group: testes derived

DKFZphtes3_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and nuclear localization signals of lyar.

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```

1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTACCTTA AAAATGGTAT
151 TTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGCGA AACTGTGAA TGCTTTTCTT GCATTGACTG
251 CGGTAAGAT TTCTGGGGCG ATGACTATAA AAACCCAGTG AATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAGGCT ATGAAGGTAA AACCCACAAA
351 GGCACATCA AACAGCAGGC GTGGATTGAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCA AAGTGAGAGA ACTTTAGAG CAAATTAGTG
451 CTTTTGACAA CGTTCCCGAG AAAAGGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTTATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTTCTGAA GCTTCCAACA GCGAACCATG CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAAGG AAAAGAGAAA
751 AGAAGAACT AAAATTAGAA AACCACAGG AAAACTCAAG GAATCAGAA
801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTTGAGG CTGGTGGGGA
851 GGAATGCCCT GAGGCCAATG GCTTCGAGG GAAGAGGAGC AAGAAGAGA
901 AGCAGCGCAA GGACAGCGCC AGTGAGGAAG AGGCACGCGT GGGCGCAGGG
951 AAGAGGAAGC CGAGGCACCT GGAAGTTGAA ACAGATTCTA AAGAAGAAA
1001 CATGAAGCTC CCAGAGCATC CTGAGGCGCG AGAACAGAA GACGATGAGG
1051 CTCTGCAAAA AGGTAAATTC AACTGGAAGG GAATATTAA AGCAATCTG
1101 AACAGGCCCC CAGACAATGA AATAACATC AAAAGCTAA GGAAGAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCAGAA TCCGAGAGG
1201 AACTCCTGGT CATCTTTAAC AAGAAAATCA GCAAGAACC TACCTTTAAG
1251 TTATTAAAGG ACAAAAGTCAA GCTTGIGAAA TGAACATTG TGIATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTTCCTCTCT TCACTGCTGT TTATAAAATG
1351 TGTAATGAAT TCTAACAAC CAAATTTTGC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAAATAT
1451 ATCTGGTCC AAACCTCAAA AAAAAAATA AAAAAAATA AAAAAAATA
1501 AAA
```

BLAST Results

No BLAST result

Medline entries

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379
 Category: strong similarity to known protein
 Classification: Cell division
 Prosite motifs: ATP_GTP_A (60-68)

```

1 MVFFTCNACG ESKKIQVEK HVSVCNCEC LSCIDCGKDF WGDDYKNHVK
51 CISEDQKYGG KGYEGKTHKG DIKQAWIQK ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNMKN SLKVHNEISL DQVWNIFSEA SNSEPVNKEQ
151 DQRLPHVPAN PHAEISTKVP ASKVKDAVEQ QGEVKNKRE RKEERQKKRK
201 REKKELKLEN HQENSRNQK KRRKKQGEAD LEAGGEEVPE ANGSAQKRSK
251 KKKQKDSAS EEARVGAGK RRRHSEVET DSKKKMKLP EHPEGGEPE
301 DEAPAKGKFN WGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHRS
351 EEELLVIFNK KISKNPTFKL LKDKVKLVK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058.18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse
 Length = 388

HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144
 Identities = 275/388 (70%), Positives = 317/388 (81%)

```

Query:      1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGDDYKNHVKCISEDQKYGG 60
            MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGDDYK+HVKCISE QKYGG
Sbjct:      1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGDDYKSHVVCISEGQKYGG 60

Query:     61 KGYEGKTHKGDQAWIQKIELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN 120
            KGYE KTHKGD KQAWIQKI+ELIK+PNVSPKVRELL+QISAFDNVP KAKFQNMKN
Sbjct:     61 KGYEAKTHKGDQAWIQKINELIKRPNVSPKVRELLQISAFDNVPIKAKFQNMKN 120

Query:    121 SLKVHNEISLDQVWNIFSEASNSEPVNKEQDQRLPHVPANPHAEIS-TKVPASKVKDAVE 179
            SLKVH++S+L+QVM+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct:    121 SLKVHSDSVLEQVWDIFSEASSSE---QDQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query:    180 QGEVKNKREKREERQKRRKKELKLENHQENSRNQKPKRRKKQGEADLEAGGEEV 239
            +Q E KKNKREKREERQK RK+EKKEKLENHQEN R QKPKRRKK QEA EA GE+
Sbjct:    177 EQTEAKKNKREKREERQKNRKKKKELKLENHQENLRGQKPKRRKKQGEAGHEAAGEDGA 236

Query:    240 EANG-----SAGKRSKKKKQKDSASEEAA----RVGAGKRRK-RHSEVETDSKKKKM 287
            + +G G+ S++ R E+ A + AGKRRK +HS E+ KKKKM
Sbjct:    237 DSGSPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRRKPKHSGAESGYKKKKM 296

Query:    288 KLPEHPEGGEPEDEAPAKGKFNWGTIKAILKQAPDNEITIKLRKKVLAQYTYVTDEH 347
            KLPE PE GE +D EAP+KGKFNWGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct:    297 KLPEQPEEGEAKDHEAPSGKFNWGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query:    348 HRSEELLVIFNKKISKNPTFKLLKDKVKLVK 379
            EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct:    357 SHHEELLAIIFNRKISRNPFTFKVLKDRVKLLK 388

```

Pedant information for DKFZphtes3_14g5, frame 3

Report for DKFZphtes3_14g5.3

[LENGTH] 379
 [MW] 43634.03
 [pI] 9.59
 [HOMOL] PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse le-122
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
 [BLOCKS] BL00603D Thymidine kinase cellular-type proteins
 [BLOCKS] BL00530C
 [PROSITE] ATP_GTP_A 1
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 18.73 %

```

SEQ  MVFFTCNAGESVKKIQVEKHVSVCRNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYGG
SEG  .....
PRD  cccccccccccccchhhhhhhheeeccccceccccccccccccceeecccccccccc

SEQ  KGYEGKTHKGDIKQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhhhcccchhhhhhhhhhhccccchhhhhhhhhhhc

SEQ  SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEISTKVPASKVKDAVEQ
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhcccchhhhhhhhhccccccccccccceccccccccchhhhh

SEQ  QGEVKNKRERKEERQKRRKREKKELKLENHQENS RNQPKRRKGQEADLEAGGEEVPE
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccccc

SEQ  ANGSAKRKSKKKQRKDSASEEEARVGAGKRRRHSEVETOSKKKKMKLPEHPEGGEPE
SEG  .....
PRD  cccccccccchhhhhhhhhcccchhhhhhhhhccccccccccccchhhhhhhcccccccccc

SEQ  DEAPAKGKFNWKGTIKAILKQAPDNEITIKKLRKKVLAQYYTVTDEHHRSEELLVIFNK
SEG  .....
PRD  cccccceehhhhhhhhhhhccccccccchhhhhhhhhhhhhhhcccchhhhhhhhhhhhh

SEQ  KISKNPFTKLLKDKVKLVK
SEG  .....
PRD  cccccchhhhhhhhhhhccc
  
```

Prosite for DKFZphtes3_14g5.3

PS00017 60->68 ATP_GTP_A PDOC00017

(No Pfam data available for DKFZphtes3_14g5.3)

DKFZphtes3_14h21

group: nucleic acid management

DKFZphtes3_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNATg

Sequenced by BMFZ

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGA CAATGTCCCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GGTGCTTGCT AGTCGGCGAA GCTCGACAGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGGATA TAGTGTCCGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
201 CCCCAGGAGG CCGTGGCGCG TGCTCAGCAG GAACGCGCGC TGTGTTTGGC
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTGCTGGT GGGTCAAAAA
301 TAAAGAATAT ACAAAGTACA ACAACACCA CAATCCAAAT AATACAAGAA
351 CAACCAAGAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCAACGAA
401 AGCAAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAAATTACA
451 ATTCAGATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
501 GGAAGCAGAC ATAACAATGT TGTGCGAGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAAT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTATA AAGAGTCCAC TGCCACAAGT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTTAATAT
701 AACGTGGGAT GACTTGAGGC ATGGGGAGAA ACGACCTATC CCCAATCCTA
751 CCTGCACATT TGATGACGCC TTTCAATGTT ATCCTGAGGT TATGGAAAAC
801 ATTAATAAGG CAGGTTTCCA AAAGCCAACA CCTATTCAGT CACAGGCATG
851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATTGTGTTAT TTAATGCCTG GATTATATCA TCTGGTCCTT
951 CAACCCAGCC TTAAGGTCA AAGGAATAGA CCCGGCATGT TAGTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAAAATAT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTCTGCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTTAGAT GTGCGCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCTCATTC AGTTCATCGC CTCGCACAAT
1351 CTTATTTGAA AGAACCAATG ATTGTCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAAGT CAGTGAAGCA AAATATAAT GTAAACCCCG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAGCAT TAGAGAACTT TAAACAGGC AAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGAG GACTGGTGT TCCATTACAA
1801 CTTTGACTAG AAATGATTCG AGCGTTGCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAT GGAAGAGAAA ATGGAAAGAC
1951 CTCAGGAGAG GCCCAAGGAG TTTCATTAA GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTTTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAATAATA
2101 GTGTTTGAAA ATATAGAAATC CAGTGTTTTA TACTTCTTTT AATAAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648
 Category: strong similarity to known protein
 Classification: Nucleic acid management
 Prosite motifs: ATP_GTP_A (286-294)
 DEAD_ATP_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTNTTT
101 IQIQEQEPES LVKIFGSKAM QTKAKAVIDN FVKKLEENYN SECGIDTAFO
151 PSVGKGGSTD NNWVAGDRPL IDWDQIREEG LKWKQTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPINPT CTFDDAFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIQVAQTCT GKTLCYLMGP
301 FIHLVLQPSL KGQRNRPGL VLTPTRELAL QVEGECCCKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDI II ATPGRLNDLQ MSNFVNLKNI TYLVLEADK
401 MLDMGFEPQI MKILLDVRPD RQTVMTSATW PHSVHRLAQS YLKEPMIVYV
451 GTLDLVAVSS VKQNIIVTTE EEKWSHMQTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQRDRKAL ENFKTGKVR I LIATDLASRG
551 LDVHDVTHYV NFDFFRNIEE YVHRIGRTGR AGRTGVSITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRPKKFKH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14h21, frame 3

TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like
 protein."; S.pombe chromosome II pl p8B7., N = 1, Score = 971, P =
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (Saccharomyces cerevisiae), N = 1,
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces
 pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid
 Y54G11A
 Length = 504

HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREGLKWQTKWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGEK 233
      D+++E W K PI ++ YK +S + + +
Sbjct: 23 DRLKDNFSWMK-----PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQIGIDLIQVAQTGTGKT 293
      IP P +F+ AF +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASIMGEIRKNGFEKPSPIQSQMWPLLSSGQDCIGVSQTGSGKT 135

Query: 294 LCYLMMPGFIHLVLQPSL-----KGQRNRPGLVLTPTRELALQVEGECCCKYSYKGLRSVC 348
      L +L+P +H+ Q + + Q+ P ,LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALLHIDAQLAQYEKNDEEQKPSFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGNRDEQIEELKKGVDI IIATPGRLNDLQMSNFVNLKNITYLVLEADKMLDMGFEP 408

```



```

          +YGGG+R EQ+E  + GV+I+IATPGRL DL      ++L ++TY+VLDEAD+MLDMGFE
Sbjct:  196 LYGGGSRPEQVEACRGGVEIATPGRLTDLSDGVISLASVTYVVLDEADRLDMGFEV 255

Query:  409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVT 468
          I +IL ++RPDR  +TSATWP V +L Y KE ++ G+LDL + SV Q
Sbjct:  256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKAAMVAVNGSLDLTSCSKSVTQFFFEV 315

Query:  469 TEEKW---SHMQTFLOQMSSTD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
          + ++      + FL +      + K+I+V K +ADHLSSD + I+ + LHG R Q
Sbjct:  316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSKVMADHLSSDFCMKGINSQGLHGGRSQS 375

Query:  525 DREKALENFKTGKVRILIATDLASRGDLVDVTHVYNDFPRNIEEYVHRIGRTGRAGR 584
          DRE +L  ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct:  376 DREMSLNMLRSGEVQILVATDLASRGIDVPDITHVLNYDFPMDEEYVHRVGRTRGRGRK 435

Query:  585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPOGRP 644
          G +++ L ND LI ILE++ Q +P++L AE++      K + R RP R
Sbjct:  436 GEAMSFLLWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KQSGRDGPRPFRN 492

Query:  645 KK 646
          K
Sbjct:  493 NK 494

```

Pedant information for DKF2phtes3_14h21, frame 3

Report for DKF2phtes3_14h21.3

```

[LENGTH]      648
[MW]           72873.51
[pI]           8.84
[HOMOL]        TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-
101
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       04.05.03 rna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-70
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
[FUNCAT]       1 genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 2e-49
[FUNCAT]       j rna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-48
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 1e-07
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 4e-96
[PIRKW]        RNA binding 3e-87
[PIRKW]        DEAD box 5e-50
[PIRKW]        transmembrane protein 4e-27
[PIRKW]        DNA binding 3e-67
[PIRKW]        recF recombination pathway 3e-10
[PIRKW]        ATP 4e-96
[PIRKW]        purine nucleotide binding 5e-50
[PIRKW]        P-loop 4e-96
[PIRKW]        hydrolase 9e-45
[PIRKW]        protein biosynthesis 5e-50
[PIRKW]        ATP binding 1e-61
[SUPFAM]       WW repeat homology 8e-88
[SUPFAM]       DEAD/H box helicase homology 4e-96
[SUPFAM]       unassigned DEAD/H box helicases 7e-87
[SUPFAM]       ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM]       ATP-dependent RNA helicase DHH1 2e-43
[SUPFAM]       recQ protein 3e-10
[SUPFAM]       Bloom's syndrome helicase 5e-07
[SUPFAM]       translation initiation factor eIF-4A 5e-50
[SUPFAM]       recQ helicase homology 3e-10
[SUPFAM]       tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE]      DEAD_ATP_HELICASE 1

```

```

[PROSITE]      ATP_GTP_A      1
[PFAM]         Helicases conserved C-terminal domain
[PFAM]         KH domain family of RNA binding proteins
[PFAM]         DEAD and DEAH box helicases
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      8.49 %

SEQ  MSHHGGAPKASTWVVASRRSSTVSRAPERPAEELNRTGPEGYSVGRGGRWGRGTSRPPEA
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNIQSTNTTTIQIQEQPESLVKIFGSKAM
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  cccccccccchhhhhccccccccccccccccccccccccccccccccccccccccchh

SEQ  QTKAKAVIDNFVKLEENYNSECGIDTAFQPSVVGKDGSTDNNVAGDRPLIDWDQIREEG
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ  LKWQKTWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGKRPINPT
SEG  .....
PRD  chhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhheeecccccccccccccc

SEQ  CTFDADFQCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQIDLIGVAQTGTGKTLCYLMPG
SEG  .....
PRD  cccccccccchhhhhhhhhhhccccccccccccccccccccccccccccccccccccccce

SEQ  FIHLVLQPSLKGQRNRPGLVLTPRELALQVEGECCKYSYKGLRSVCVYGGGNRDEQIE
SEG  .....
PRD  eeeeeccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  ELKKGVDIIATPGRLNDLQMSNFVNLKNITYLVLEADKMLDMGFEPQIMKILLDVRPD
SEG  .....
PRD  hhhheeeccccccccchhhhhhhccccccccccccccccccccccccchhhhhhhhhhhcc

SEQ  RQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
SEG  .....
PRD  ceeeeccccchhhhhhhhhhhhhheeeccccccccccccccccccccchhhhhhhhhhh

SEQ  LQSMSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDRKALENFKTGKVRI
SEG  .....
PRD  hhhccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhccce

SEQ  LIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS
SEG  .....
PRD  eehhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhhhh

SEQ  ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERFQGRPKKFH
SEG  .....
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphtes3_14h21.3

HMM_NAME	DEAD and DEAH box helicases	
HMM	*gLpPWILRNIyeMGFEkPTPIQQaIPiLeGRDVMAQAQTGSGKTAAF	
Query	248	QCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQIDLIGVAQTGTGKTLCY 296
HMM	L+P ++H+ +P +++ Q+ P +L+L+PTRELA+Q++ EC K+++ +	
Query	297	LMPGFIHLVLQP-SLKGQRNRPGLVLTPRELALQVEGECCKYSYK-G- 343
HMM	IRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDRIeM	
Query	344	LRVSVYVGGGNRDEQIEELKGV-DIIATPGRLNDLQMSNFVNLKNITY 392
HMM	LVMDEADRLDMGFIDQIRrIMrQIPmwnRQTMFSATMPdeIqELARr	
Query	393	LVLDEADKMLDMGFEPQIMKILLDVR--PDRQTVMTSATWPHSVHRLAQS 440

```

HMM          FMRNPiRInId.MdElTtnEnIkQwYiyVerEMWkfdcLcrLie*
+++++P  + ++ D  +++ +KQ +I+  E++K +  ++++
Query      441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-EEEKWSHMQTFLQ      482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
          + + ++++G++IG+GGS I++I++ ++++I I++E+ +  + + I
Query      71  CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQEQ-P---ESLVKIF      115

HMM          G*
          G
Query      116 G      116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWlkn1...GlrvmYIHGdMpQeERdeIMddFNnGEynVLicTD
          + +++ L+ +  +I+V ++HGD++Q++R+++++F++G+ ++LI+TD
Query      497 KAVADHLSSDLILGNISVESLHGDRQDRREKALENFKTGKVRILIATD      545

HMM          VggRGIDIPdVNHVINYDMPWNPEqYIQRIgRTgRIG*
          +++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query      546 LASRGLDVHdVTHVYNFDfPRNIEEYVHRIGRTGRAG      582

```

DKF2phtes3_14p14

group: testes derived

DKF2phtes3_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```

1 GAAGCCCAAG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTT
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTG
101 GGTGCTGAAG GCCAGGCAGA GCATTGGCT GTAGGGAGGC CGATCTCTCT
151 CGGCGCTGTT ACCGGCGGCT CTTTGTCTT AGACCTGGGG TTCTTGGCCT
201 CACGGATTCC AAGGAATGGA ACCTTGGGCC ATGCGTGTGA ACGAGCTCTA
251 TGTGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGTCTGA
301 ACATCAGTTT ACCCAATCTG CACTGCCGAGT TGGTTGGGCT TGACATTTCAG
351 GATGAGATGG GCAGGACAGA AGTGGGCCAC ATCGACAAC CTATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGCCGCTT CGAGGGGGCAG TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTTCTT
501 GCTGTCTCCC CGCTCCCTGG AAACCTGGTTG TGGAGGCACT CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TCGGTGTCCA GCTGCCTGGC
601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTTCAG CCCCCAGCCC TGGGTTCAAG
751 TCCAGCTCT ACCCCTTCTT GGGCCCTACA AGTCACTTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
851 AGTTTGTGTG GGGATTAAT GAGCTAATGC AGATTTCATC ATTCAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CAGCGTCTCT GGGCTGGTGG GGCTCCCAT TCGGTAGAGG GAGACAGTCT
1001 ACAAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAAG
1051 CAAGGGGAGT GAGATGAGT GGGCTTGAAG GTACCTTGTC CGCTCAGAAG
1101 GACCAATCAA GGTTCAGTGT TGTTTGTGCC TCAGAACCAAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACGCTGTGAG GCAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGCCACAGAG GGTAAAGTGT CTTGCCCGAG
1351 GTCACACAGC TATCTGCATG TCCCACAAC CTCCCTCCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCTCT CAGACCTTGA CACTGAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTTGTGTTG TCCAGTATAG
1551 CATTTAACTG CATCAGAAC AGTATGAAAA GACCAGGAAT CCAGATTTCCT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAG TGGGCAGTGT TCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCCTGG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGTCTAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGCT
1851 GGGGAAATC TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACCAGCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCTGT
1951 GTTCACTGGG GGCTCTCTCT GCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCCGTA ATTTGGATT TATGGTGGTT GTGATGAAA GCCATTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGTAACTAGA TTTATGCTT AGAATTGGAC
2101 TATCCAATAG GTTGCCACCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTAC AATTCACTCA CTCATTCTCA TCAACCACAT TTCAAGTGCT
2201 CAAAGCCAGC TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGCACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCACGAGG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCTCTACT AAAAATACAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCGATGAGCC
2551 AAGATCTGAT CACTGCACTC CAGTCTGGGC GACAGAGTGA GACTCATCT
2601 CAAACAAAC AAAAAGGA TGGGGCTGGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GAGGATAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGG CACATGTGGA

```

```

2751 AGCAAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCTGTGTT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACTG GTCATTCCAC TCTGTTTCC TTTGGGGTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAAAGTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTCTCTG ACCACAGATG TTGAGGTGGG AGTCATTAGT GAATAGATGA
3001 TGTGGAAGT CATGGGTCCT CAGAGTGGGG GCTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTGTCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTCC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG
3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCAGCAC TTTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCGAGCTG GTCAACATGT
3601 CAAAACCCCT TATCTACAAA AAAATACAAA AGTTAACCAG GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGCTCT CTGAGCCAG
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAACCCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTTA AATATTTTAA TTACTATTTA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159
 Category: putative protein
 Classification: no clue

```

1 MERWAMRVNE LYVDDPKDGS GKKIDVSLNI SLPNLHCELV GLDIQDEMGR
51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPLSPFY LLPFAVSPL
101 PGNWLWRHSL DLTLTQPPAS EGSCPAAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSLQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p14, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3_14p14, frame 3

Report for DKFZphtes3_14p14.3

```

{LENGTH}      159
{MW}           17778.55
{pI}           5.74
{FUNCAT}      99 unclassified proteins      [S. cerevisiae, YAL042w] 5e-04
{KW}           Alpha_Beta

```

```

SEQ  MERWAMRVNELYVDDPKDGSKKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM
PRD  ccchhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  KIPLNNGAGCRFEGQFSINKVWKPLSPFYLLPFAVSPLPGNWLWRHSLDLTLTQPPAS
PRD  eeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

WO 01/12659

PCT/IB00/01496

SEQ EGSCPAAWPFLRLRIWMGVQAPWGFKPLMAGSGRSYSSLQ
PRD cccccchhhhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3_14p14.3)

(No Pfam data available for DKFZphtes3_14p14.3)

DKFZphtes3_14p7

group: testes derived

DKFZphtes3_14p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1  GGAATCCAAA GAAACAGTTA TGATGGGGGA CTCTATGGTG AAAATAAATG
51 GGATTATTTT AACAAAATCA AATGCTATTT GCCACTTAAA GAGTCACCCA
101 CTTCAGCTAA CTGATGATGG AGGCTTCAGT GAAATAAAGG AGCAAGAAAT
151 GTTCAAAGGA ACAACATCTT TACCATCTCA TCTCAAGAAT GGAGGGGACC
201 AGGGGAAGAG ACATGCGAGG GCCTCATCAT GCCCCAGTAG CTCAGACCTG
251 AGCAGGCTGC AAACCAAAGC AGTCCCAAAA GCTGACCTGC AAGAAGAGGA
301 CGCAGAAATA GAAGTAGACG AAGTCTTTTG GAATACAAGG ATTTGACCGA
351 TTTTGCCTGA ATTAGAAAAG GAAGAAAACA TTGAAACGGT TTGTGCTGCT
401 TGCACACAACTTCATCATGC TTTAGAGGAA GGAAACATGC TTGGAATAAA
451 ATTTAAGGGA AGAAGTATTC TCCTGAAGAC CCTGTGTAAG CTAGTTGATG
501 TTGGTTCAGA CTCGCTCAGC CTTAAACTTG CAAAAATAAT TCTAGCACTT
551 AAAGTGAGTA GAAAGAATCT TCTTAATGTC TGCAAACTTA TATTTAAAT
601 TAGCAGGAAT GAGAAGAATG ATTTCTTGAT TCAAAATGAC AGCATTCTGG
651 AATCATTATT GGAGGTACTA AGAAGTGAAAG ACCTGCAAAAC TAACATGGAA
701 GCTTTTATAT ACTGTATGGG GTCTATAAAG TTCATTCTG GAAATCTGGG
751 ATTTCTTAAT GAAATGATCA GCAAAGGTGC TGTGGAATA CTGATAAATT
801 TGATAAAACA AATAAATGAG AACATCAAGA AATGGGTAC ATTTTGGCT
851 AATTGCGGCC ACTTGCTAGT CCAGGTGACT GCTACATTGA GAAACTTGGT
901 TGATTCACTA TTAGTAAGAA GTAAGTTCCT AAACATCAGT GCCCTTCCCC
951 AGCTCTGCAC GGCAATGGAA CAGTACAAGG GTGACAAGGA CGTCTGTACC
1001 AATATTGCCA GAATATTCAG CAAACTTACT TCTTACCGTG ACTGCTGCAC
1051 AGCCTTGGCC AGCTATTCCA GATGTTATGC CTTATTCTG AATCTAATTA
1101 ACAAACTCCA GAAGAAGCAG GATTTAGTCG TCCGTGTTGT TTTTATCTTT
1151 GGCAACCTGA CGGCAAAAAA TAACCAGGCT CGTGAACAA TTTCCAAAAG
1201 GAAAGGGAGC ATCCAAACTC TGCTGTCATT ATTCCAGACG TTCCATCAGC
1251 TGGATCTGCA TTCCAGAAG CCGGTGGGCC AACGAGGCGA GCAGCACAGG
1301 GCGCAGAGGC CGCCGTCAGA GGCAGAGGAC GTGCTCATCA AGCTGACTCG
1351 TGTGCTGGCC AACATTGCCA TCCACCGGGG CGTGGGCCCG GTGCTGGCCG
1401 CCAACCCGGG GATAGTGGG CTGCTCCTGA CCACGCTGGA ATACAAGTCA
1451 CTTGATGATT GTGAGGAGCT GGTGATCAAT GCTACAGCGA CAATCAACAA
1501 TTTATCTTAC TACCAAGTGA AGAATTCAT AATTCAAGAC AAAAAGCTAT
1551 ATATTGCTGA ATTGCTCTTA AAGCTTCTTG TCAGTAACAA CATGGATGGA
1601 ATCCTGGAGG CTGTGCGTGT TTTGCGAAAT CTCTCCAGG ACCATGATGT
1651 CTGCGATTTC ATTGTGCAGA ACAATGTCCA CAGGTTTCATG ATGGCGCTGC
1701 TGGATGCTCA GCATCAGGAT ATCTGCTTTT CTGCTGTGG TGTCTCTCTC
1751 AATCTCACTG TGGATAAAGA CAAGCGTGTG ATCTTGAAG AAGGAGGTGG
1801 CATTAAAAAG TTAGTGGACT GTTTAAGAGA TTTGGTCTCT ACTGATTGGC
1851 AGCTGGCCTG CTTGGTTTGT AAAACTTTAT GGAACCTCAG TGAAAAACATC
1901 ACTAATGCTT CGTCATGTTT TGGAAATGAA GACACCAACA CACTCTTACT
1951 CTTGCTCTCA TCATTTTGTG ATGAAGAACT AGCACTGGAT GGCAGTTTGT
2001 ATCCAGACCT AAAAACTAT CACAACTCC ATTGGGAAAC AGAATTCAAA
2051 CCTGTGGCAC AGCAGCTTCT AAACCGAATT CAGAGACATC ACACCTTCTC
2101 GGAACCCCTG CCCATTCCTT CTTCTAACA TGATGCAGAT TAACAGTAGA
2151 AACGAGAACT CACGCTCTCC TCATTCTTAA GAACGTGTAA CAACGCTGAA
2201 CATTTTTTTC AGCATTAACA AATGTGGAAA GTTTTCAAG AACTGGTTTT
2251 AGTGAGTAGC TGAAGTATTT TTTAAATTA AGCATTCTCT CTTGTAGGT
2301 ATTATGAAA AATGAATATA CACATTATAT TTCCTGTTGA GAGAAATGTA
2351 AGATGAAAT ATGTGCATT TCAAGTAAAT GACTTTTCT TCTATTCTCT
2401 ATTAACAAT TTAGTCTAG TCTTAAAAA AAAAAAATA AAAAAAATA
2451 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702
Category: putative protein

```
1 MMGDSMVKIN GIYLTNSNAI CHLKSHPLQL TDDGGFSEIK EQEMFKGTTS
51 LPSHLKNGGD QGKRHRARASS CPSSSDLSRL QTKAVPKADL QEEDAIEIVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IPKISRNEKN
201 DSIQNDISIL ESLLLELRSE DLQTNMEAFI YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIQINENIK KCGTFLPNSG HLLVQVTATL RNLDVSSSLVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RYALFLNLI NKYQKKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQI
401 LLSLFQTFHQ LDLSHSQKPVG QRGEQHRARQ PPSEAEVLI KLTRVLANIA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSYYQV
501 KNSIIQDKKL YIAELLKLL VSNMMDGILE AVRFGNLSQ DHDVCDIFVQ
551 NNVRFRMMAL LDAQHQDIFC SACGVLLNLT VDKDKRVILK EGGGKIKLVD
601 CLRDLGPTDW QLACLVCCTL WNFSENITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFKPVA QQLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_14p7, frame 2

TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,
complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete
cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 45/163 (27%), Positives = 77/163 (47%)

Query: 442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSYYQVK 501
L +++ NI+ H G P VG L + S D+ EE VI T+ NL+ +

Sbjct: 483 LMKMIRNISQHDG--PTKNLFIDYVGDAAQI---SSDEEEEFVIECLGTLANLTIPDL 537

Query: 502 -NSIIQDKKLYIAELLKLLVSNMMDG-ILEAVRVFGNLSQDHDVCDIFVQNNVHRFMA 559
++++ KL + L KL D +LE V + G +S D + ++ + ++

Sbjct: 538 WELVLKEYKL-VPFLKDKLPGAAEDDLVLEVIMIGTVSMDDSCAALLAKSGIIPALIE 596

Query: 560 LLDAQHQDIFCSACGVLL--NLTVDKDKR-VILKEGGGIKLVDCLRD 604

LL+AQ +D F C ++ + + R VI+KE L+D + D

Sbjct: 597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIKETQAPAYLIDLMD 644

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 42/178 (23%), Positives = 82/178 (46%)

Query: 169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDISILESLLLEVLRSDELQTNME 227
K K L V ++ LL V L+ ++ + + + ++N +I+ L++ L + N E

Sbjct: 263 KTFKRYQGLVVKQEQLLRVALYLLNLAEDTRTELKMRNKNIVHMLVKALDLD---NFE 318

Query: 228 AFLYCMGSIKFISGNLGFLENEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVT 287
+ + +K +S + N+M+ VE L+ +I +E++ L + +

Sbjct: 319 LLLVVSFLKKSIFMENKNMDVEMDIVEKLVKMIPEHEDL-----LNITLR 366

Query: 288 ATLRLNVDSLSVRSKFLNISALPQLCTAM--EQYKGDKDVCT--NIARI--FSKLTSYRD 341
L D+ L R+K + + LP+L + E YK +C +I+ F + +Y D

Sbjct: 367 LLLNLSFDTGL-RNKMVQVGLLPRLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424

Query: 342 CCTAL 346
C L
Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAE L L L K L L V S N M D G I L E A V R V F G N L S Q D H D V C D F I V Q N N V H R F M M L L D A Q H Q D I C F S 571
I +L+K L +N+ ++ V LS + + +V+ ++ ++ ++ +H+D+
Sbjct: 304 IVHMLV KALDRDNFELLILVVSFLK KLSIFMENKNDVMEDIVEKLVKMI PCEHEDLLNI 363

Query: 572 ACGVLLNLTVDKDRVILKEGGG I K K L V D C L R D L G P T D W - Q L A C L V C K T L W N F S E N I T N A 630
+LLNL+ D R + + G + K L L G ++ Q+A +C L++ S +
Sbjct: 364 TLRLLNL S F D T G L R N K M V Q V G L L P K L T A L L --- G N E N Y K Q I A -- M C - V L Y H I S M D - D R F 416

Query: 631 S S C F G N E D T - N T L L L L S S F L D E E L A L D 657
S F D L+ +L DE + L+
Sbjct: 417 K S M F A Y T D C I P Q L M K M L F E C S D E R I D L E 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 L I F K I S R N E K N - D S L I Q N D S I L E S L L E V L R S E ----- D L Q T N M E A F L Y C M G S I K F I S G 241
L I +++R N N + L+ N++ L +L V L R + +L T N+ +C S G
Sbjct: 155 L I L Q L A R N P D N L E E L L N E T A L G A L A R V L R E D W K Q S V E L A T N I I Y I F F C S S F S H F H G 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 L V Q V T A T L R N L --- V D S S L V R S K F L N I S A L P Q L C T A M E Q Y K G D K V C T N I A R I F S K L T S 338
+++ T L N L +D L V + + +P L + +D+ +C S
Sbjct: 521 V I E C L G T L A N L T I P D L W E L V L K E Y --- K L V P F L K D K L P G A E D D L V L E V V - I M I G T V S 576

Query: 339 Y R D C C T A L A S Y S R C Y A L F L N L I N K Y Q K K Q D L V R V V F I L G N L T A K N N Q A R E Q F S K E K G S I 398
D C A L + S + L + N Q + + V +++++ + R + K E +
Sbjct: 577 M D D S C A A L L A K S G I I P A L I E L L N A Q Q E D E F V C Q I I Y V Y Q M V F - H Q A T R D V I I K E T Q A P 635

Query: 399 Q T L L S L 404
L + L
Sbjct: 636 A Y L I D L 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 C E - E L V I N A T A T I N - N L S Y Y Q - V K N S I I Q D K K L Y I A E L L L K L L V S N M D G I L E A V R V F G N 537
C E ++N T + N L S+ ++N ++Q + + L L L + N I A + V +
Sbjct: 355 C E H E D L L N I T L R L L L N L S F D T G L R N K M V Q --- V G L L P K L T A L L G N E N Y K Q I -- A M C V L Y H 409

Query: 538 L S Q D H D V C D - F I V Q N N V H R F M M L L D A Q H Q D I C F S A C G V L L N L T V D K D R V I L K E G G G I K 596
+S D F + + + M L + + I + N L +K ++ E G +K
Sbjct: 410 I S M D R F K S M F A Y T D C I P Q L M K M L F E C S D E R I D L E L I S F C I N L A A N K R N V Q L I C E G N G L K 469

Query: 597 K L V D C L R D L G P T D W Q L A C L V C K T L W N F S E N I T N A S S C F G N E D T N T L L L L S S F L D E E L A L 656
L+ R L D L+ K + N S++ F + L +SS +EE +
Sbjct: 470 M L M K -- R A L K L K D --- P L L M K M I R N I S Q H D G P T K N L F - I D Y V G D L A A Q I S S D E E E E F V I 522

Query: 657 D 657
+
Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 L N I S A L P Q L C T A M - E Q Y K G D K V C T N I A R I F S K L T S Y R D C C T A L A S Y S R C Y A L F L N L I N K 362
L N +A L L + E +K ++ T N I I F +S+ + Y + A L +N+I+
Sbjct: 171 L N E T A L G A L A R V L R E D W K Q S V E L A T N I I Y I F F C S S F S H F H G L I T H Y - K I G A L C M N I I D H 229

Query: 363 Y Q K K Q D L 369
K + +L
Sbjct: 230 E L A R H E L 236

Pedant information for DKFZphtes3_14p7, frame 2

Report for DKFZphtes3_14p7.2

[LENGTH] 708
[MW] 79266.35
[pI] 6.57

```

(FUNCAT)      30.25 vacuolar and lysosomal organization [S. cerevisiae, YEL013w] 3e-04
(FUNCAT)      06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w]
3e-04
(FUNCAT)      09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YEL013w] 3e-04
(BLOCKS)      BL00923F Aspartate and glutamate racemases proteins
(BLOCKS)      BL00288B Tissue inhibitors of metalloproteinases proteins
(PSITE)        MYRISTYL 9
(PSITE)        AMIDATION 1
(PSITE)        CK2_PHOSPHO_SITE 12
(PSITE)        PKC_PHOSPHO_SITE 7
(PSITE)        ASN_GLYCOSYLATION 11
(KW)           Alpha_Beta
(KW)           LOW_COMPLEXITY 7.49 %

```

```

SEQ      ESKETVMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFKGTTSLSH
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LKNGGDQGRKHARASSCPSSDLSRLQTKAVPKADLQEDAEIEVDEFWNTRIVPILRE
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LEKEENIETVCAACTQLHHALEEGNMLGNFKGRSILLKTLCKLVVDGSDLSLKLAKII
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LALKVSRKNLLNVCKLIFKISRNEKNDLSIQNDSILESLEVLSEDLQTNMEAFLYCMG
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SIKFISGNLGFNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLVQVATLRNLV
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      DSSLVRSKFLNISALPQLCTAMEQYKGDQDVCTNIARIFSKLTSYRDCCTALASYSRCYA
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LFLNLINKYQKQDLVVRVVFILGNLTAKNNQAREQFSKERGSIQTLLSLFQTFHQLDLH
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SQKPVGQRGEQHRAPPPSEADVLIKLTRLVLANIAIHGPGVPLAANPGIVGLLLTTLE
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      YKSLDDCEELVINATATINLSYYQVKNISIIQDKLYIAELLKLLVSNMMDGILEAVRV
SEG      .....
PRD      hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      FGNLSQDHDVCDFIQNNVHRFMALLDAQHQCICFSACGVLLNLTVDKDKRVILKEGGG
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      IKKLVDCLRDLPDQLACLVCVCKTLWNFSENITNASSCFGNEDTNTLLLSFLDEEL
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      ALDGSFDPDLKNYHKLHWETEFKPAQQLNRIQRHHTFLEPLPIPSF
SEG      .....
PRD      hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_14p7.2

PS00001	206->210	ASN_GLYCOSYLATION	PDOC00001
PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00001	311->315	ASN_GLYCOSYLATION	PDOC00001
PS00001	385->389	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00001	500->504	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	584->588	ASN_GLYCOSYLATION	PDOC00001
PS00001	628->632	ASN_GLYCOSYLATION	PDOC00001
PS00001	632->636	ASN_GLYCOSYLATION	PDOC00001
PS00001	635->639	ASN_GLYCOSYLATION	PDOC00001
PS00005	173->176	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005

WO 01/12659

PCT/IB00/01496

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_l4p7.2)

DKF2phtes3_15a13

group: testes derived

DKF2phtes3_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST

hits

S.cerevisiae Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAGCGCA  TCGCGTCGG  GCACAGCGG  TGCAGCCTG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCGAG  TTGCAGAGGA  CTCCCATGAG
151  TGCACCTGGTA  TTTCCCAATA  AGATATCAAC  TGAACACCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTTGAGG
251  GGAATATTCC  CAGAATGCGC  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGTCAAAATA  CTGAGAGAA  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAACCCAG  AAGATCCTCA  GACAATTCCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTTCTAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTTG  TCTACTGACA  CCAAGAAAGC  AAGCATTCTC
551  CTCATTGCGA  AGATTTATAT  CCTAATGCAA  AATCTGGGGC  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGTTATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAAA
801  TTGACTCAAC  TATACTATCA  CCAAAACAAA  TAAAAACACC  ATTTCAAAAA
851  ATCCTGAGGG  ACAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTTGGAG  ATTGAACATA  AAATGGAAGA  ACAGGAAAAA  AACCTGCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAAATTATG
1001  AGGCTCTAA  AAAGTCCAGA  TCTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAATAGT
1201  CCTCCATCAC  TTTGATTCTT  CTAGTCAAGA  GTCAAGTCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTTCGAGAGT  TCTTCTCACC  ATTTAACTG  AAGGACCCTA  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACATAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTC
1501  GGTAATAAGT  AAAATTTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCCTTTAAA  TATATACTAA  CTGTTAAGGC  CCTCTAATGC  CATTTTTCTA
1601  AACAGTAATG  TTTACTTTGG  TATTAAAAAT  TGGTATGGAT  TCACCTTTTA
1651  CTTATGTTAA  AATTATACCA  TTTAACTGGC  TCTTTGTGCA  TTGTCGTGTT
1701  ATTAACACAA  TGTCTCTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAGG
1801  GCGGCGCGT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387
Category: similarity to known protein

```
1 MATAQLQRTF MSALVFPNKI STEHQSLVLV KRLLAVSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDK NCPGSQLVK WMLGCDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNNGLP MDFISKQSN ESSMLSTDTK KASILLIRKI
151 YILMQNLGPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERENID STILSPKQIK TPFQKILRDK
251 DVEDEQEYHT SDDLDIETKM EEQKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLSSHSQV EQLVKNKSEL DMSKSTRSG KVFQNMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSSS QESVPKRRKF SEPKEHI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15a13, frame 2

TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.
Length = 562

HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22
Identities = 84/290 (28%), Positives = 145/290 (50%)

```
Query: 22 TEHQSLVLVKRLLAVSVSCITYLRGIFPECAYGTRYLDDLCVKILREDRNCPGSQLVKW 81
      TE SL+L + LL +++ I+Y+RG+FPE + + + L+KI + S+L+ W
Sbjct: 11 TEQDSLLLRNLLRIAFNISYIRGLFPEKYFNKSVPALDMKIKLMPMDAESRRLIDW 70

Query: 82 M-LGCDYDALQKKYVYT-----NPEDPQTISECYQFKFKYTNNGP--LMDFISK--NQSN 130
      M G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct: 71 MEKGVYDALQRKYLKTLMFISICETVDGPMIEE-YSFSFSYSDSDSQVMMNINRTGNKKN 129

Query: 131 ESSMLST-----DTKKASILLIRKIYILMQNLGLPNDVCLTMKLFYYDEVTPPDYQPP 184
      ST + ++ ++R+ LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct: 130 GGIFNSTADITPNQMRSSACKMVRTLVLQMLRTLDKMPDERTIVMKLLYYDOVTPPDYEP 189

Query: 185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVT-----ERERENIDSTILS 235
      F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct: 190 FFRGCTEAEQYVWTKNPLRMEIGNVNSKHLVLTLVKVSVDPCEDENDMQD-DGKSIG 248

Query: 236 PKQIKTPFKILRDKDVEDEQEYH-----TSDDLDIETKMEEQKNPASSE 281
      P + Q D ++ QE+ DD D E ++ ++PA +E
Sbjct: 249 PDSVHDD-QPSDSSEISQTQENQFIVAPVEKQDDDDGEVEDDNTQDPAENE 300
```

Pedant information for DKFZphtes3_15a13, frame 2

Report for DKFZphtes3_15a13.2

```
[LENGTH] 387
[MW] 44417.64
[pI] 5.57
[HOMOL] TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 03.13 meiosis [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIL072w] 7e-11
[PIRKW] nucleus 2e-09
[PIRKW] zinc finger 2e-09
```

```

[PIRKW]      DNA binding 2e-09
[PROSITE]    MYRISTYL      1
[PROSITE]    CAMP_PHOSPHO_SITE      3
[PROSITE]    CK2_PHOSPHO_SITE      12
[PROSITE]    PKC_PHOSPHO_SITE      7
[PROSITE]    ASN_GLYCOSYLATION      3
[KW]         Alpha_Beta

SEQ  MATAQLQRTPMASLVFPNKISTEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDD
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LCVKILREDKNCPCGSQLVWMLGQYDALQKKYVYTNPEDPQTISECYQFKKYTNNGPL
PRD  hhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  MDFISKQSNESMLSTDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPD
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  YQPPGFKDGDCEGVIFEQEPMYLVNGEVSTPFHIFKVKVTTTEREREMIDSTILSPKQIK
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TPFQKILRDKDVEDEQEYHTSDDLOICTKMEEQEKNPASSELEEPSLVCEDEIMRSKES
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  PDLISHSQVEQLVNKTSLEDMSESKTRSGKVFQNMANGNQPVKSSKENRKRKSQHESGR
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  IVLHHFDSSSQESVPKRKRFSEPKHEI
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosite for DKF2phtes3_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phtes3_15a13.2)

DKF2phtes3_15c24

group: metabolism

DKF2phtes3_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxyypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxyypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatgG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
1 CGAAGGCGGC GGCAGAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGAAC TGCCAGGAG AGGAGTCTGC AGGTCCCAG GAGCGGCGAC
151 GGAGGGGCGC GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAAT GGAATTGTAA
251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTGAAG TACACAACATA
501 TAATATAACC ACAGTGGAAA ACTTTCAACA TTTCATGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTTGA ATGAAGTTGG
651 ACAAAACATG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGCCTG TGCTCCACCA
751 CTTGTAGTTG CTGCAATAT TGATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTGTGCAGCC AGTCTTCTTA CCACATAGGG TGTGGTTGCT GGGATCTTAG
851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCACTG CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAAGG GAACGTAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATA CAAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGGAAGA TTCTGGTGAA AGCTTGGGAAG ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATAATT TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAACT TGATAAAACT
1351 TAGGGCAACA TTAATTAAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAATCCTG TGACTTGCCCT GTTTCTCCCC GCTCCAACGA AATCATTAACT
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTCTATAAG AAAATTGCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACCTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAAAATAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAA
1951 AAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404
Category: similarity to unknown protein
Classification: Metabolism
Prosites motifs: D_2_HYDROXYACID_DH_1 (76-105)

```
1 MAESVERLQQ RVQELERELA QERSLQVPRS GDGGGGRVRI EKMSSEVVDS
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LLFDYDKVEL ANMNRLLFFQ HQAGLSKVQA AEHTLRNINP DVLFEVHNYN
151 ITTVENFQHF MDRISNGGLE EGKPVDLVLS CVDNFEARMT INTACNELGQ
201 TWMESGVSEN AVSGHIQLII PGESACFACA PPLVVAANID ETKLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPMTSMKPN
301 PQCDNRNCRK QQEYKKKVA ALPKQEVQEE EEEIHEDNE WGIELVSEVS
351 EELKNFSGP VPDLPPEGITV AYTIPKKQED SVTELTVEDS GESLEDLMAR
401 MKNM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c24, frame 1

TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid
T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1
YAC yUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus
fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796.4 gene: "moeB"; product: "MoeB"; Staphylococcus
carneus molybdenum cofactor biosynthetic gene cluster, complete
sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.
Length = 419

HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122
Identities = 241/367 (65%), Positives = 293/367 (79%)

```
Query: 37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96
      R +IEK+S+EVVDSNPYSRLNAL+RMGIV++YE+IR VA+VGVGGVGSV AEMLTRCG
Sbjct: 48 RQKIEKLSAEVVDSNPYSRLNALQRMGIVNEYERIREKTVAVGVGGVGSVVAEMLTRCG 107

Query: 97 IGKLLLFDDYDKVELANMNRLLFFQHQAGLSKVQA AEHTLRNINPDVLFVHNYNITTVEN 156
      IGKL+LFDDYDKVE+ANMNRLLF+Q+QAGLSKV+AA TL ++NPDV EVHN+NITT++N
Sbjct: 108 IGKLLLFDDYDKVEIANMNRLLFYQPNQAGLSKVEAARDTLIHVNPVQIEVHNFNITMDN 167

Query: 157 FQHFMDRISNGGLEEGKPVDLVLSVDNFEARMTINTACNELGQTWMESGVSENAVSGHI 216
      F F++RI G L +GK +DLVLSVDNFEARM+N ACNE Q WMESGVSENAVSGHI
Sbjct: 168 FDTFVNRIRKGSLLTDGK-IDLVLSVDNFEARMVNMVMAENQIWMESGVSENAVSGHI 226

Query: 217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLKFLNLF 276
      Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTTM VVAG LV N LK+LLNF
Sbjct: 227 QYIEPGKTACFACVPLVVASGIDERTLKRKGVCASLPTTMVAVAGFLVMNTLKYLNF 286

Query: 277 GTVSFYLGYNAMQDFFPMTSMKPNPQCDNRNCRKQQEYKKKVAALPKQ-EV-IQEEEEEI 334
      G VS Y+GYNA+ DFFP S+KPNP CDD +C ++Q+EY++KVA P EV +EEE +
Sbjct: 287 GEVSQYVGYNALSDFFPDSDIKPNPYCDDSHCLQRQKEYEEKVANQFPVDLEVEVPPEETV 346

Query: 335 IHEDNEWGIELVSEVSEELKNFSGPVPDLPPEGITVAYTIIPKKQEDSVTELTVEDSGESL 394
      +HEDNEWGIELV+E SE + S + G+ AY P K+ D+ TEL+ + +
Sbjct: 347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAYE-PIKR-DAQTELSPAQA--AT 399

Query: 395 EDLMKMKMN 403
```


DKFZphtes3_15c6

group: transmembrane protein

DKFZphtes3_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```
1 GAGACACTGA GCCCGAGAC AGTGAGTGGT GGCTCACTG CTCTGCCCCG
51 CACCCGTGCA CCTCCACTTT GCCTTGTGG AAGTGACCCA GCCCCTCCC
101 CTTCCTTCT CCCACCTGTT CCCCAGGACT CACCCAGCC CCGCCTGCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGGCTTGGAG CCAGAGACAC
251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGCCGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCGTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCTGGCTA GGAGTGTTC
401 TGTTTCCTCC TTTGCCTACC ACTCTGGGGT GGGCCAGTGT GTGGGAAGC
451 TGGCTGTCGG ATGGTAGCTA TTCCACCTC TGCTGCCTG CCGCCTGCT
501 GTCTGGGCA TGCTGCAGTA CCTGTGCCA GGATTGGTTT TAAATTGTGA
551 AATAATTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGAAGTCTC
601 TCCACAGCC TGCGCTTGCC TCCCTGCCTC ATCTCTATT TCATTCCACT
651 ATGCCCCAAG CCCTGGTGGT CTGGCCCTTT CTTTTCTCT CTATCCTCAG
701 GGACCTGTGC TGCTCTGCCC TCATGTCCCA CTGGGTGTT TAGTTGAGGC
751 ACTTTATAAT TTTTCTCTG TCTGTGTTT CTTTCTGCT TATTCCCTG
801 CTGTGTCCTG TCCTTAGCAG CTCACCCCA TCCTTGCCA GCTCCTCCTA
851 TCCCGTGGG ACTGGCCAAG CTTTAGGGAG GCTCCTGGT TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGTG GGTGAGGCC GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACTTC ACTGCATCCT TGCCCATTC AGCCCGGCT
1001 TTCATGATGC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGTGA TCCAAGTGA GTGGGACCCC CTAAGAGGT CAGGAAGTGG
1151 ACACTAACAT CTGTGAGGT GTTGACTTGA AAAATAAAG GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGCGGCC GCTCTAGAGG
1251 ATCAAAGCTT ACGTAAAAAA AAAAAAATAA AAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118
Category: putative protein

```
1 MVAIPPSACL PACCPGHGAV PVPRIGFKFV NNFPFGLVDV NRAREVLPTA
51 CACLPASSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHPV LGCLVEALYN
101 FSLVLCFLL YFPAVSCP
```

No BLASTP hits available

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

HSPs:

Query: 15 PGHGAVPVPRIGKFVNFFPFGLVDVNRAREVLPTACACLPASSLFSFHYAFSPGGLALS 74
PG GA P+ R+ F+ PF + +E+ A C P SSL+ A G L
Sbjct: 52 PGRGA-PLARVTRFH----PFR---KQKELFVAEEVCTPVSSLYCGKKATLVGNV 103

Query: 75 FSSYPQPGVLLCP---HV-PLGCLVEALYNFSLVL 105
S P+G V+ C HV G L A +++V+
Sbjct: 104 LRSIPEGAVV-CNVEHHVGDGVLARASGDYAI 137

Report for DKFZphtes3_15c6.2

```
(LENGTH)      118
(MW)           12413.79
(pI)           7.53
(PROSITE)     LEUCINE_ZIPPER 1
(PROSITE)     MYRISTYL       1
(PROSITE)     ASN_GLYCOSYLATION 1
(KW)          TRANSMEMBRANE 1
```

```
SEQ      MVAIPPSACLPAACCPGHGAVPVPRIGRIFVNNFPGFLVDNVRAREVLPTACALPASSLF
PRD      cccccccccccccccccccccccceeecccceehehhhhhcccccecccccccc
MEM

SEQ      SFHYAPSPGGGLALSFSYYPQGPVLLCPHVPGLCLVEALYNFSLVLCFLLYFPVASC
PRD      eeccccccccceeeccccccccccccccccchhhhhhhccccchhhhhhhcccccccc
MEM      .MMMMMMMMMMMMMMM.
```

Prosites for DKFZphtes3_15c6.2

PS00001	100->104	ASN GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_15c6.2)

DKF2phtes3_15g14

group: testes derived

DKF2phtes3_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```
1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATGT TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTAGAAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACTT CCAATGTGA AGAAAAAGCT
401 GATGTTTTAA GCTCCTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCCTGT GATGTAAGAG AGAAGTGGCT TTCTAAAAA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAGAGG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCTTTT TAGTAACGTG
601 AGGAAAAAAC AGTGAATTTG TTGTAAACCC AAATCTTGAA TATAAAGAAC
651 TTTGTCAATT GGTATCTGAA GAGGAAGCAT TTGACTTTT TAAATATTTG
701 GATGCAAGA AAGAAAAATC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCACAGA AAGCTGTGCC ACCATTTTGT CACAAAAAG TTTGGAACCC
801 TTTGGGAAC CAAATCTTTT TCTAAATGA ATTGCAATGC TGGTAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCCGGGA AAAGCACACA AACGTGGGAA
901 AAGGCCCTCT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTACGAAA GGAACACCTG GAAATGTTTG AAGCGATTGG TTTTITAGCT
1001 ATCAAACTTG GTGTTATTCC TTCCGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCATATC AAGCAATGGT TGTAGAAAA GTGACTCCAG
1101 AGAGGTTGAA AATATTGAA AAGAAATTG AAAAGAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATTCCTGT AGACTTGGTC AGCTCAAAAGG
1201 AATACACTTT GATATTGTCA TTAGAAATTT AAAAAACAA ATAAATGATT
1251 CTGCAACCTT GAGGAGAGA ATTATGGAAG CAATAGAAAA TGTTAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAAATT GACTAGCTTT GCTGAAGAA GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAG ACTTGGATGA TCCTGTAAT
1451 AGAGCAAGA AGTATTTCTT TCAAACTGAG GATGCTAAAG GCACACTTTC
1501 ATTGATGCGT GAATTCAAAG TGGCTGAGAG AGCATTTGTT GAGGCATTGC
1551 ACCGCTTTGG CATGACCGAG GAAGGTTGTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTCCA TGCGCATATT CTATGTTTAC GCATATACCA GCAAAATTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCACTACC
1851 CGAAGACAA AGTAGGGCAG TGGTACCATG ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCAACCTTT CCTGTTTTTG AAATATTGTA
2201 TCAGAACAA ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTATAT CTGGCATAC GCTGACCCCT
2351 TTGACCAATT GTAATTTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCAAT TCTTAGATTG TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTGTAAAA TTTAATTTT TGTGTTGTTA
2601 ACTATCCAC ATTATAATC ATCCTTACC ATTTATATC AGTTAAATAT
2651 GGGTGTGTT GGGAGGAATG ACTGCCATG AGACATGTAT TGATTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAATTTAA
```

```

2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAAG
2801 CAGTTTAAAC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTG CGTAACTAAA
2951 GTAAGTAAAA TGATGTGTAT CAACCTTAGT GTAAATATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTCACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAA AGCACAGATA TTGAGACAAA
3151 ACTAACCAG TTTGAACCCG GGCACCTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAGTTAT TTAACCTCAT GGGCTTCAGT TTCAACATCT GTAAAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAAT TTTCTAAGTT
3301 AATATATGTA AAGCAACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAAGCAA TAAATGAAAA
3401 TCCCTTCTGT TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701
 Category: similarity to unknown protein

```

1 MEEDTDYRIR FSSLCFFNDH VGFHGTIKSS PSDFIVIEID EQQLVNKTI
51 DEPIFKISEI QLEPNFPKK PKLDLQNLSL EDGRNOEVHT LIKYTDGQDN
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNNF ACDVREKWL
151 KTELIGLPPE FSIGRILOKN QRASLHSAIR QKFPFLVTVG KNSIVVKPN
201 LEYKELCHLV SEEEAFDFEK YLDAKKENSK FTFFKPDNTKD HRKAVHHFVN
251 KKFGLNVETK SFSKMNCAG NPNVVTVRF REKAHKGKR PLSECQEGKV
301 IYTAFTLRKE NLEMFPAIGF LAIKLGIVPS DFSYAGLKDK KAITYQAMV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 QINDSANLR ERIMEAIENV KKKGFVNYG PQRFGKGRKV HTDQIGLALL
451 KNEMMKAIKL FLTPEDLDDP VNRAKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHFRGM TEEGCIQAMF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDDIDEN FPNKSIHLVT EEEGSANMYA IHQVVLPLVG
601 YNIQYKKNKV GQWYHDILSR DGLQTCRFKV PTLALNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTKGSHI DETALSLIIS FDLBASCYAT VCLKEIMKHD
701 V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15g14, frame 2

TREMBL:SPBC1A45P_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid clA4 left hand region 1-26184 bp
 Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)
 Length = 676

HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
 Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSIIVVKPNLEYKELCHLVSEEEAFDFEK-YLDAKKENSKFTFFKPDNTKDHRAVHHFV 249
 + E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPRLR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNLVETKSF SKMNC SAGNPV VTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307
+ F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNSRTNKQEKINQTRDANGVENWGYGPKDFIHFTL 224

Query: 308 RKENLEMF EAIGFLA I KLGVIPSD-FSYAGLKDKKAITYQAMVVRKVTPERLK NIEKEIE 366
KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVISIKIGLDRNLALNRTL- 282

Query: 367 KKRNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKKG FV 426
K M + N D SL LG LKGN F +VIR++ N +L E + +++ + GF+

Sbjct: 283 -KGMIIGNYNFSDASLNLGDLKGNFV VVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGPRFGKGRKVHTDQIGLALLKNEMMKA I KFLT PEDLDDPVNR-AKKYFLQTEDAK 485
NY+G QRF G + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFGT F-SISTHTIGRELLSNWKKAAELILSDQDNVLPKSKEARKIWAETKDAA 399

Query: 486 GTLSLMPEFKVRERALLEALHFRFGMTEEGCIQ--AWFS----LPHSMRIFYVHAYTSKI W 539
L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPROCLAENALLYSLSNQRKEEDGTYSENAYTAIMKIPRNLRTMYVHAYQSYVW 459

Query: 540 NEAVSYRLETYGARVVGDLV-----LDEDIDDENFPNS-----KIH LVTEEEGS 585
N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NSIASKRIELHGLKLVVGDV LVIDTSEKSP LISGIDDED FDEVDREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVVLFPVLGYNIQYPKNK-VGQWYHDILSRDGLQTCRFKVP TLKLNIPGCGYRQI 644
+ Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLFPSPGFDVLYPSNEELKQLYVDILKADNMDFNMRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671
++ P +L Y+++ D + + + +D

Sbjct: 580 IQPKPSLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01
Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFIVIEIDEQQQLVNKTI DEPIFKISEIQLEPNFNNPKPKLDLQNL SLE 81
GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIK YTDGDQNHQSGS--EKEDTI-VDGTSKCEEKADVLSSFLDEKTHELLN 138
R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKI H 161

Query: 139 NFACDVREKWL SKTELIGLPPE-FSIGRILDKNQRASLSAIRQ 181
+RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698
++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKFLGTSAYATMALREL MK 660

Pedant information for DKF2phtes3_15g14, frame 2

Report for DKF2phtes3_15g14.2

{LENGTH} 701
{MW} 80700.96
{PI} 7.31
{HOMOL} PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae) 2e-51
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53
{BLOCKS} BL01268C
{BLOCKS} BL01268B
{BLOCKS} BL01268A
{SUPFAM} hypothetical protein HI0701 3e-06
{PROSITE} MYRISTYL 7
{PROSITE} AMIDATION 2
{PROSITE} CAMP_PHOSPHO_SITE 1
{PROSITE} CK2_PHOSPHO_SITE 16
{PROSITE} TYR_PHOSPHO_SITE 1
{PROSITE} PKC_PHOSPHO_SITE 13
{PROSITE} ASN_GLYCOSYLATION 5
{KW} Alpha_Beta

```

SEQ MEEDTDYRIRFSSLCFFNDHVGFGHTIKSSPSDFIVIEIDEQGLVNKTIDEPIFKISEI
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QLEPNFPKKPKLDLQNLSELDGRNQEVHTLIKYTDGDNHQSGSEKEDTIVDGTSKCEE
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ KADVLSSFLDEKTHELLNNFACDVREKWLSTELIGLPPEFSIGRILDKNQASLHSAIR
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEAFFFKYLDAKKENSÁFTFKPDNKKD
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HRKAVHHFVNKKFGNLVETKSFSKMNCASAGNPVVTVRFREKAHKGKRLSECEGKV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ IYTAFTLRKENLEMFAGFLAIGLVIKGLVIPSDFSYAGLKDKKAITYQAMVVRKVTPERLKN
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ IEKEIEKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KKKGFVNYGPRFGKGRKVHTDQIGLALLKNEMMKAIKLFLTPEDLDDPVNRKKYFLQ
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TEDAKGTLMLPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRIFYVHAYTSKIWN
PRD hcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EAVSYRLETYGARVVGQDGLVCLDEDIPDENFPNSKIHLVTEEGSANMYAIHQVVLVPLG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ YNIQYPKNKVGQWYHDILSRDGLQTCRFKVPVLKLNIPGCYRQILKHPCLNSYQLMEDHD
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ IDVKTGSHIDETALLISFDLDASCYATVCLKEIMKHDV
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15g14.2)

DKFZphtes3_15h1

group: testes derived

DKFZphtes3_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACGAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
51 TCTCTAAATC CGGTCAACAT GTCGGACCCC GAAGGCGAGA CCTTGCAGAG
101 CACCTTTCCC TCTTATATGG CCGAAGGCGA GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCTCGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTTCTG TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCCAG GAAGCCATCA
451 ACAACTCAGT GGAAGTCTCT TCTTCCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAA ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATG AACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGA
601 AGGCTCTGCT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGAGCTC
651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGGCG GCCTGACTGT GGAGGACCTC ATCATGACCG
751 GCATCAACTA CTTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTCGG
851 GGACCACAAA CGCCGTCCCT CACAGACAGC CCATTACATC CTCAGAGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGTCTT
951 CAGAAAAGCTG AGAAAAGTCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAC AAGGATGAAC TGGTTGAAA CTTGTATAGC TGCATAGGGA
1051 ATGCCCATGAT TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAAGATCC CTCTGGCAAA AACCACCTG
1251 GAGAAAGACCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCTGGG CAGGCCCAGA ATTATGGCGA GAAGTCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGCACAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGATATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAAA GAAAAAAGCG AGGGAGAAGC
1601 TTCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCAGAGAAG GTGGTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAGGGG CTGTCAAGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CTTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAAG.AAC TTTCAGAAAT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAATTT GGAGAAATAG GAGAAACGAA
2051 AAAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTTC AAGCGATTGG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACTCTT
2251 GCAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672
Category: similarity to known protein

```
1 MSDPEGETLR STFPSYMAEG ERLYLCEFS KAAQSFNSAL YLQDGDKNCL
51 VARSKCFLKM GDLSRLKDA EASLQSDPAF CKGILQKAET LVTMGDFEFA
101 LVFYHRGYKL RPDREFRVGI QKAQEAINNVS VGSPSSIKLE NKGDLNFLSK
151 QAENIKAAQKQ PQPMKHLHP TKGEPRKAS LKSEKTVRQL LGELYVDKEY
201 LEKLLDDEL IKGTMKGLT VEDLIMTGIN YLDTHSNFWR QQKPIYARER
251 DRKLMQEKWL RDHKRRPSQT AHYILKSLED IDMLTSGSA EGSQKAEKV
301 LKKVLEWKE EVPNKDELVG NLYSCIGNAQ IELQMEAL QSHRDLLEIA
351 KEYDLPAKS RALDNIGRVF ARVGKFFQAI DTWEEKIPLA KTTLEKTWLF
401 HEIGRCYLEL DQAWQAQNYG EKSQQCAEEE GDIEWQLNAS VLVQAQVKL
451 RDFSAVNNF EKALERAALV HNNEAQAII SALDDANKGI IRELKRTNYV
501 ENLEKSEGE ASLYEDRIIT REKDMRRVRD EPEKVVQWD HSEDEKETDE
551 DDEAFGEALQ SPASGQSVE AGKARSDLGA VAKGLSGELG TRSGETGRKL
601 LEAGRESRE IYRRPSGELE QRLSGEFSRQ EPEELKKLSE VGRREPEELG
651 KTQFGEIGET KKTGNEMEKE YE
```

BLASTP hits

Entry AF039202.1 from database TREMBL:
product: "Hsp70/Hsp90 organizing protein"; *Cricetulus griseus*
Hsp70/Hsp90 organizing protein mRNA, complete cds.
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782.1 from database TREMBL:
product: "myosin heavy chain"; *Argopecten irradians* myosin heavy chain
mRNA, complete cds.
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:
stress-induced protein stil - soybean
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3_15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15h1, frame 3

Report for DKFZphtes3_15h1.3

```
{LENGTH}      672
{MW}           76655.61
{pI}           5.49
{HOMOL}        PIR:S56658 stress-induced protein stil - soybean 6e-10
{SUPFAM}       tetratricopeptide repeat homology 1e-07
{PROSITE}      MYRISTYL 7
{PROSITE}      AMIDATION 3
{PROSITE}      CAMP_PHOSPHO_SITE 4
{PROSITE}      CK2_PHOSPHO_SITE 15
{PROSITE}      TYR_PHOSPHO_SITE 1
{PROSITE}      PKC_PHOSPHO_SITE 11
{PROSITE}      ASN_GLYCOSYLATION 2
{KW}           All_Alpha
{KW}           LOW_COMPLEXITY 4.76 %
```

```
SEQ  MSDPEGETLRSTFPSYMAEGERLYLCGEFSKAAQSFNSALYLQDGDKNCLVARSKCFLKM
SEG  .....
PRD  ccccccccccceccccccccccccccccchhhhhhhhhhhhhhhccccceehhhhhhhhhhh

SEQ  GDLSRLKDAEASLQSDPAFCGILQKAETLVTMGDFEFALVFYHRGYKLKRPDREFRVGI
SEG  .....
PRD  hcchhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhh
```

[illegible]

Prosite for DKFZphtes3_15h1.3

PS000001	128->132	ASN_GLYCOSYLATION	PDCC000001
PS000001	438->442	ASN_GLYCOSYLATION	PDCC000001
PS000004	265->269	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	605->609	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	613->617	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	636->640	CAMP_PHOSPHO_SITE	PDCC000004
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	66->69	PKC_PHOSPHO_SITE	PDCC000005
PS000005	136->139	PKC_PHOSPHO_SITE	PDCC000005
PS000005	180->183	PKC_PHOSPHO_SITE	PDCC000005
PS000005	183->186	PKC_PHOSPHO_SITE	PDCC000005
PS000005	186->189	PKC_PHOSPHO_SITE	PDCC000005
PS000005	214->217	PKC_PHOSPHO_SITE	PDCC000005
PS000005	342->345	PKC_PHOSPHO_SITE	PDCC000005
PS000005	564->567	PKC_PHOSPHO_SITE	PDCC000005
PS000005	596->599	PKC_PHOSPHO_SITE	PDCC000005
PS000005	660->663	PKC_PHOSPHO_SITE	PDCC000005
PS000006	2->6	K2_PHOSPHO_SITE	PDCC000006
PS000006	66->70	K2_PHOSPHO_SITE	PDCC000006
PS000006	93->97	K2_PHOSPHO_SITE	PDCC000006
PS000006	171->175	K2_PHOSPHO_SITE	PDCC000006
PS000006	220->224	K2_PHOSPHO_SITE	PDCC000006
PS000006	277->281	K2_PHOSPHO_SITE	PDCC000006
PS000006	382->386	K2_PHOSPHO_SITE	PDCC000006
PS000006	392->396	K2_PHOSPHO_SITE	PDCC000006
PS000006	481->485	K2_PHOSPHO_SITE	PDCC000006
PS000006	507->511	K2_PHOSPHO_SITE	PDCC000006
PS000006	512->516	K2_PHOSPHO_SITE	PDCC000006
PS000006	542->546	K2_PHOSPHO_SITE	PDCC000006
PS000006	548->552	K2_PHOSPHO_SITE	PDCC000006
PS000006	628->632	K2_PHOSPHO_SITE	PDCC000006
PS000006	663->667	K2_PHOSPHO_SITE	PDCC000006
PS000007	506->515	TYR_PHOSPHO_SITE	PDCC000007
PS000008	119->125	MYRISTYL	PDCC000008
PS000008	132->138	MYRISTYL	PDCC000008
PS000008	213->219	MYRISTYL	PDCC000008

WO 01/12659

PCT/IB00/01496

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15h1.3)

DKFZphtes3_15i5

group: cell structure and motility

DKFZphtes3_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)
"radial spokehead" part of flagella in *Chlamydomonas*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```
1 CACCTCGGCC CGCTCCCGGC GCCCTCCACG GGTAAACGGCC CCCTCTCTCG
51 GTGCTCAGAA ACCGGCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101 TCTGCAAGCC TTTCTCCTAG CATCTCTGAG CTCCTGGGGA ACCATGGGAG
151 ACCTGCGGCC CTACCTGAG CGCCTGCCCC AGCAGCCTCC GGGCCGGAGG
201 ACTTCTCAGG CCTCCACAG GCGGCACAGT CGGACCAAG CTCAGGCCCT
251 GGCAGCGGAC CCGAGGAGA GGCAGCAGAT ACCTCCAGC GCCACCGTAA
301 AGCGCCCTGG TTGGTCACAG AGGGGACAGC TGTCCTCAAC GGAGAACTTG
351 CTGATGGCCCC AGGTCTTCCA GCGTGAGGAA GCCCGGCTGG CTGGCATGGA
401 GTACCCATCT GTGAACAGCG GCTTTCCTTC AGAGTTCAG CTCAGCCCTT
451 ACTCTGATGA AAGCAGGATG CAGGTCCGCG AGCTCACCAC CAGCCTAATG
501 CTGCAAGCGG TCCAGCAGGG CCAAGCAGC CTGTTCAGC AACTGGACCC
551 CACCTTCCAG GAGCCCCCAG TCACCCCTTT GGGCCAGTTT AACTCTTACC
601 AGACAGACCA GTTCTCTGAA GGTGCCCTTT GGGCCAGTTT AACTCTTACC
651 GACCCCTGCC TTCACTTCTT GCGCTCTGAG CTGGGCTTCC CACACTACAG
701 TGCCCAAGGT CCGAGCCCGG AGCCTCTGGA GCTGGCCGTG CAGAACGCCA
751 AGGCCCTACT GCTGCAGACC AGCATCAATT GCGACCTCAG CCGTACGAGG
801 CACCTGGTAA ATCTGCTGAC CAAGATCCTG AACCAGCGCG CTGAGGACCC
851 CTGTCTGTCT CTGGAGTCTC TCAACCGCAC CACCGAGTGG GAGTGGTTCG
901 ACCCCAGCTT GGACACGCTG CCGGACGACC CCGAGATGCA GCCCACCTAC
951 AAGATGGCGG AGAAGACGAA GCGGCTGTTC ACCCGGAGTG GAGGCGGCAC
1001 TGAAGGCGAA CAGGAGATGG AGGAGGAGGT GGGGGAGACA CCAATGGCCA
1051 ACATCATGGA GACTGCTTTC TACTTCGAGC AGGCCGCGT CGGCTGAGC
1101 TCGGACGAGA GCTTCCGATC TTTCTGGGCC ATGAAACAGC TGGTGGAGCA
1151 GCAGCCCATC CACACCTCTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACCT GGTGGCCGAG GTGGAATTCC GGGAGGGGGA GGAGGAGGCA
1251 GAGGAGGAGG AGCTGGAGGA GATGACGGAA GTGGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGCGG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCCTAAGTC CGTATGGAAG CCGCCGCCCC TGATCCCAAC GGAGGAGAGC
1401 CGCTCAGGCG CCAACAACTA CCGTACTTTT GTGTCAACG AGCCGGCCCT
1451 GCCATGGAGC CGGCTGCCCC ACCTCACTCC AGCCAGATG GTGACGCCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACAGGCC AGTCGTGAGC
1551 TACCCACCCCT TCCCGGGCAA CGAGGCCAAC TACCTGCGGG CCCAGATAGC
1601 CCGCATCTCG GCGGCCACGC AGGTCAAGCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAGG GTGCTGCTGG GCGGACTCC
1701 TACGAGGAGA ACCCGGACTT CGAGGGCATC CCCGTGCTGG AGCTGGTCTGA
1751 CTCCATGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGGAGGGCG
1801 GCTGCATCTG GGTGAACCCCT TTGCAGAGA CAGAGGAGGA GAGGAGCCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG GAAATCATGC
1901 GGTGGGCCCC CCACTGCTAA CGCCACTTTC AGAAGATGGA GAAATCATGC
1951 ACCTGGCACCC CTGGACCAAC CGCCTGTCTT GCAGCCTCTG CCCGAGTAC
2001 TCAGTGGCGG TTGTGCGCTC CAACCTCTGG CCCGGGGGCT ATGCGTATGC
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT ACAAGAGTAC
2101 GCCCGGAGAG CTTCACACCG GCCCTGCCAG CCGCCATTCA ACAAGAGTAC
2151 CCCAGTGGCC CAGAGATCAT GGAGATGAGT GACCCACAG TGGAGAGGGA
2201 CAGAGGCTCT AAAGCAGCCC CCGGAGGAGC CCGGAGGAGG ACAGAGGAGG
2251 AGGAGGAGGG CGAGGAGGAG GAGGAGGGCG AGGAGACAGA TGACTGAGGC
```

```
2301 CCACCCCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCTCTTA
2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCTCAGA
2401 GGGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAAATAAA TTCCTCCACG
2451 GCATTAAAAA AAAAAAAGG AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

86251010:
Molecular cloning and expression of flagellar radial spoke and dynein
genes of
Chlamydomona

81142496:
Radial spokes of Chlamydomonas flagella: polypeptide composition and
phosphorylation of
stalk components.

9450971:
Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm
axonemes: involvement of the protein in the regulation of sperm motility.

Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717
Category: strong similarity to known protein

```
1 MGDLPYPYPER PAQPPGRRRT SQASQRRHSR DQAQALAADP EERQIIPDA
51 QRNAPGWSQR GSLSQQENLL MPQVFQAEAA RLGGMEYPSV NTGFFPSEFQP
101 QPYSDESRMQ VAELTTSLML QRLQQGQSSL FQQLDPTFQE PPVNPFLGQFM
151 LYOTDQFSEG AQHGPIYRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ
201 NAKAYLLQTS INCOLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTTQWE
251 WFHPKLQTLR DDPEMQPTYK MAEKQKALFT RSGGGTEGEQ EMEEVGETP
301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KQLVEQQPIH TCRFMGKILG
351 IKRSYLVAEV EFREGEEAE EEEVEEMTEG GEVMEAHGEE EGEDEEEKAV
401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV
451 NARKIKKFFT GYLDPVVSYS PFFPGNEANY LRAQIARISA ATQVSPILGFY
501 QFSEEEGDEE EGGAGRDSY EENPOFEGIP VLELVDSMAN VVHHTQHILP
551 QGRCTWVNPL QKTEEEEDLG EEEKADGEP EEVEQEVGPP LLTFLSEDAE
601 IMHLAPWTTT LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH
651 KYSPESFNPA LPAPIQQEYP SGPEIMEMSD PTVEEEQALK AAQEALGAT
701 EEEEEGEEEE EGEETDD
```

BLASTP hits

Entry U73123.1 from database TREMBL:
product: "radial spokehead"; Strongylocentrotus purpuratus radial
spokehead mRNA, complete cds.
Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:
radial spoke protein 6 - Chlamydomonas reinhardtii
Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15i5, frame 3

Report for DKFZphtes3_15i5.3

[LENGTH] 717
[MW] 80913.61
[pI] 4.36

```

(HOMOL)      TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus
radial spokehead mRNA, complete cds. 1e-130
[PROSITE]    TRANSFERRIN_1      1
[PROSITE]    MYRISTYL           5
[PROSITE]    AMIDATION          2
[PROSITE]    CAMP_PHOSPHO_SITE  2
[PROSITE]    CK2_PHOSPHO_SITE   14
[PROSITE]    TYR_PHOSPHO_SITE   1
[PROSITE]    GLYCOSAMINOGLYCAN  1
[PROSITE]    PKC_PHOSPHO_SITE   8
[PROSITE]    ASN_GLYCOSYLATION  1
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY     21.48 %

```

```

SEQ  MGDLPYPYPERPAQPPGRRTSQASQRRHSRDQAQALAADPEERQQIPDQARNAPGWSQR
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ  GSLSQENLLMPQVFQAEEARLGGMEYPSVNTGFPSEFQPPQPYSDSRMQVAELTTSML
SEG  .....XXXXXXXXXXXXX.....
PRD  cccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  QRLQGGSSLFQQLDPTFQEPVFNPLGQFNLYQTQDFSEGAQHGPYIRDDPALQFLPSEL
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPELEAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNRPEDPLSVL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccchhh

SEQ  ESLNRTTQWEFHPKLDLRRDPEMQPTYKMAEKQKALFTRSGGGTEGEQEEMEEVGETP
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhchhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhcccc

SEQ  VPNIMETAFYFEQAGVGLSSDESFRILAMKQLVEQQPIHTCRFWGKILGIKRSLVAEV
PRD  .....
PRD  cccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhccccchhhhh

SEQ  EFREGEEEEEVEEMTEGGEVMEAHGEEEGEEDEEKAVDIVPKSVWKPVPPIPKESR
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhheeecccccccccccccccccccc

SEQ  SGANKLYFVCNEPGLPWTRLPHVTPAQIVNARKIKKFTGYLDPVVSYPFPFGNEANY
SEG  .....
PRD  cccceeeeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  LRAQIARISAATQVSPGLGFYQFSEEEGDEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhhhh

SEQ  WVHTQHILPQGRCTWVNPQLKTEEEEDLGEEEKADEGPVEEVEQVGPPLTPLSEDAE
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhccccccccccccccccchhhhhhhhhhhccccchhhhhcccccccccccccccccccccccc

SEQ  IMHLAPWTRLSCSLCPQYSVAVVRSLNWPAYAYASGKKFENIYIGWGHKYSPEFNP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LPAPIQEQYPSGPEIMEMSDPTVEEQALKAQEQALGATEEEEEEEEEEEGEETDD
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Prosite for DKF2phtes3_1515.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

WO 01/12659

PCT/IB00/01496

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKF2phtes3_15i5.3)

DKF2phtes3_15j18

group: testes derived

DKF2phtes3_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GSF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGCGGGCCCG TGTCCCGCC TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCATCTCAGT GGAATGCCCT ACCCCAGTCC TGGGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACCCC CTCATCCCTG CACCCCTTCC CTGCCCTGGA
451 TTCTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGCTCTCTTT TGCTTTCCAG CCAGTGTITT GGGGTTTGAA
601 GTTGGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTTCAT
751 CAGGTGATTC TTGTTGTAG CAAAGTTTAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTGG TGAACCTCTA AAAAAAAAAA
851 AAAAAAAAAA GGGCGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPAPWRD VLQQPGVGGE GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHPFSEGV GTQVECLTPV LRLESDMART
101 AHPSSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAECS WSQPGAGS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKF2phtes3_15j18, frame 2

Report for DKFZphtes3_15j18.2

```

(LENGTH)      148
(MW)           15665.78
(pI)           8.91
(PROSITE)      MYRISTYL      3
(CK2_PHOSPHO_SITE) 1
(KW)           Irregular

SEQ  MFGCPVRCPKPPTQLISGEASAARLPARVDVLQPGVGGEGGLRISWQGAPKSRVRPAFI
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPVPFTVLQSQHYHPFSEGVGTQVECLTFVLRLLESDMARTAPHPSSLHPPFPAWDSSSPVH
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  CGAPLPSAHGGFPARAEGSWSQPGAGS
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_15j18.2

```

PS00006      82->86  CK2_PHOSPHO_SITE      PDOC00006
PS00008      38->44  MYRISTYL              PDOC00008
PS00008      42->48  MYRISTYL              PDOC00008
PS00008      49->55  MYRISTYL              PDOC00008

```

(No Pfam data available for DKFZphtes3_15j18.2)

DKF2phtes3_15j3

group: nucleic acid management

DKF2phtes3_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276C, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits
YGR276c = ribonuclease H
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp

Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CTTCTTTGCG CAGGCAGACG
151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGGAAAGGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAAG CAGGCAGGCC
251 CCAATAAAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCCGAG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTACTGACAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
401 TATGCAGTTC TGGGCAATC CAATGTTCCT AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACCACTAA ACAACGTAGT GGTTTTTGTT CTGCAGGGAA
501 TGAGTCAGCT ACACTTTAC AGGTCTCTAT TGGAGTTTGG ATGCTTTCGA
551 AAAGCATTTA GACATAAAT CCGCTTGCTT CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAACCTGAACA AAGAGCTGGA GATCTGCCCA
651 AGACAATGGA AGGGCCCTTA CCTTCTAATG CAAAAGCCGC CATCAACCTT
701 CAGGATGATC CCATCATTCA AAAGTATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCCCT CTGACAAAGG AGGAAATGAG AACGTTTCAC TTTCCATTAC
801 AAGGTTTTCC TGATTGTGAA AACTTTTTAC TTACCAATG TAATGGTTCT
851 ATAGCAGACA ATAGTCCTCT CTTTGGACTT GACTGTGAAA TGTGCCTCAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGGTTGCT GAAGGAGGCT
951 GCTGTGTTAT GGATGAAGT GTCAAACTCG AAAACAAGAT TCTGGACTAC
1001 CTCACCAGCT TTTCCGGGAA CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTGGG ATCTCAGAGC ACTGAAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAGT TCTTAGCCAA AGTTATTTTG GGAAGGATA
1251 TACAGTGTCG AGACAGACTT GGTCTGATG CCACAGAAGA TGCTAGAACA
1301 ATCCTTGAAT TGGCTCGGTA TTTCTTTAAG CATGGCCCAA AAAAGATTGC
1351 AGAATAAAT CTAGAAGCAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTC AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGCTTTTC AAATAAAGAG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGCCCT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACTGTCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTAAGAGCTT TGGCCGAGTC CAGTCAATGA CTTTGTGTTCT
1801 TGAACCCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCAGC CTTGATTGTG
1851 ACACCCCTCG GAATGAGCTG GAAGGAGATT CTGAAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTTT CTGGCCCTGG AAGCTGTGAT CTTGCCATAA GATCTTAAAA
2001 GTGGAAAGCA GAAAAAATAC TGTTTCCTGA AATTCAAAG TTTGGCAGT
2051 GCCCAGCAGG CCTCAACAT TCTCACAGGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCA GGCACCTCCA TGCCTGGCTC AGAGGCTTAC
2151 CACCTGAATC AACAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTGCAGAC TCTGAAACTG GACCACCCGA AGATAGCAGC
2251 CTGGCGCTGG AGCCGGAAGA TTTGAAAGCT CTACACACAG TTGTGCCCGG
2301 GCACTCTCTG CCTCATCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGGAAATAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCGTG TGTTCGTGAG TCGGCCTGCC ATGTTTCCAT GTGCCATTTC
```

```

2451 TTACCCCTTG TAGGCAATGG CAAAGAAATGT GGTACAGCTG TAGCCTCCCC
2501 AACCAGCAGA CAGTTTTATG GAAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA
2651 AGGCGCGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743
 Category: similarity to known protein

```

1 MEPEREGTER HPRKVRESRQ APNKLVGAAE AMKAGWDL EE SQPEAKKARL
51 STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LDCMCLTSK GRELTRISLV
251 AEGGCCVMEDE LVPKENKILD YLTSFSGITK KILNPVTTKL KDVRQLKAL
301 LPPDAVLVGH SLDLDRALK MIHPYVIDTS LLYVREQGRR FKLKFLAKVI
351 LGKDIQCPDR LGHDATEDAR TLELARYFL KHGPKKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLDCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQKK YCFLKFKSFG SAQQALNILT GKDWKLGRRH ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQALQTLK LDHPKIAAWR WSRKIGKLYN
701 SLCPTGLCLI LLPGTKSTHG SLSGGLMGI KEEESAGPG LCS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15j3, frame 2

TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product";
 Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,
 N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid
 C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637_9 gene: "SPAC637.09"; product: "putative
 exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P
 = 2.8e-27

>TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo
 sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.
 Length = 547

HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 358/373 (95%), Positives = 358/373 (95%)

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN 164
 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN
 Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN 60

Query: 165 AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD 224
 AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD
 Sbjct: 61 AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD 120

Query:	225	NSPLFGLDCEM-----CLTSKGRLETRISLVAEGGCCVMDELVKPENKIL	269
		NSPLFGLDCEMCLTSKGRLETRISLVAEGGCCVMDELVKPENKIL	
Sbjct:	121	NSPLFGLDCEMARTTTFNSIGVLAECCLTSKGRLETRISLVAEGGCCVMDELVKPENKIL	180
Query:	270	DYLTFSFGSITKKILNPVTTTKLKDVRQLKALLPPDAVLGVHSLDLDLRALKMIHPYVIDT	329
		DYLTFSFGSITKKILNPVTTTKLKDVRQLKALLPPDAVLGVHSLDLDLRALKMIHPYVIDT	
Sbjct:	181	DYLTFSFGSITKKILNPVTTTKLKDVRQLKALLPPDAVLGVHSLDLDLRALKMIHPYVIDT	240
Query:	330	SLLYVREQRRRFLKFLAKVLGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE	389
		SLLYVREQRRRFLKFLAKVLGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE	
Sbjct:	241	SLLYVREQRRRFLKFLAKVLGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE	300
Query:	390	LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAAGLPSSR	449
		LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAAGLPSSR	
Sbjct:	301	LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAAGLPSSR	360
Query:	450	NCQTIKCLSNKEV 462	
		NCQTIKCLSNKEV	
Sbjct:	361	NCQTIKCLSNKEV 373	
<p>Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284 Identities = 175/179 (97%), Positives = 177/179 (98%)</p>			
Query:	538	LETRQVRQVPVTELTLDCCDTLVNELEGDSNGSGIYLSGVSETFKEQLLQEPRLFGLGLEAV	597
		L++VQRVPVTELTLDCCDTLVNELEGDSNGSGIYLSGVSETFKEQLLQEPRLFGLGLEAV	
Sbjct:	368	LSNKEVRQVPVTELTLDCCDTLVNELEGDSNGSGIYLSGVSETFKEQLLQEPRLFGLGLEAV	427
Query:	538	ILPKDLKSGKQKKYCFLKFKSFGSAQAQALNLTGKDWKLGKRHALTPRHLHAWRLGLPPE	657
		ILPKDLKSGKQKKYCFLKFKSFGSAQAQALNLTGKDWKLGKRHALTPRHLHAWRLGLPPE	
Sbjct:	428	ILPKDLKSGKQKKYCFLKFKSFGSAQAQALNLTGKDWKLGKRHALTPRHLHAWRLGLPPE	487
Query:	658	STRPLGLRVVPPFPEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK	716
		STRPLGLRVVPPFPEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK	
Sbjct:	488	STRPLGLRVVPPFPEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK	546

Pedant information for DKFZphtes3_15j3, frame 2

Report for DKFZphtes3 15j3.2

```

[LENGTH]      743
[MW]           83536.58
[PI]           8.87
[HMOL]        TREMBL:AC004381.4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
Chromosome 16 BAC clone CIT9875K-44M2, complete sequence. 0.0
[FUNCAT]      01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
[FUNCAT]      99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
[FUNCAT]      05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 1e-10
[FUNCAT]      04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 1e-10
[FUNCAT]      03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
[PROSITE]     MYRISTYL 5
[PROSITE]     AMIDATION 1
[PROSITE]     CK2 PHOSPHO_SITE 8
[PROSITE]     TYR PHOSPHO_SITE 1
[PROSITE]     GLYCOSAMINOGLYCAN 1
[PROSITE]     PKC PHOSPHO_SITE 16
[PROSITE]     ASN GLYCOSYLATION 2
[PFAM]        RNA recognition motif. (aka RRM, RBD, or RNP domain)
[KW]          Alpha Beta

```

SEQ MEPEREGTERHPRKVRRESRQAPNKLVGAAEAMKAGWOLEESQPEAKKARLSTILFTDNCE
PRD cccchhhhhccccchhhhhhhhhcchhhhhhhhhccccccccchhhhhhhccccccccc

SEQ VTHDQLCELLKYAVLKGKSNVPKPSWCQLFHQNHLNNVVVFVLQGMSQLHFYRFYLEFGCL
PRD eehhhhhhhhhhhhhhhccccccccceeeccccccceeeeeecchhhhhhhhhhhhhhh

SEQ RKAFRHKFRLPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSNAKAAINLQDDPIIQKY
PRD hhhhhhhhhccccccccchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhccccccc

SEQ GSKKVG LTRCLLTKEEMRTFHFPLQGFDPDCENFLLTKNGSIADNSPLFGLDC EMC L TSK
PRD ccccccc h h h h h h h h h h h h h h h h c c c c c c c c c c e e e e c c c c c c c c c c e e e e c c c c c c c c c c

SEQ GRELTRISLVAEGGCCVMDELVKPENKILDYLTFSFGITKKILNPVTTKLKDVQRQLKAL
PRD cchhhhhheeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhh

```

SEQ  LPPDAVLVGHSLDLDLRALKMIHPYVIDTSLLYVREQGRFRLKFLAKVILGKDIQCPDR
PRD  hccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhcccccc

SEQ  LGHDATEDARTILELARYFLKHGPKKIAELNLEALANHQEIQAGQEPKNTAEVLQHPNT
PRD  cccccchhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhcccccccccccc

SEQ  SVLECLDSVGQKLLFLTRETDAAGLPSSRNCQTIKCLSNKEVLEQARVEIPLFPFSIVQF
PRD  cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccc

SEQ  SFKAFSPVLTEEMNKRMRKWTETISTVYAGPFSKNCNLRLKRLFSFGPVQSMFTVLET
PRD  eeeeeehhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhcccccccc

SEQ  RQVQRPVTELTLDCLTLVNELEGDSENGQSIYLSGVSETFKEQLLQEPRLFLGLEAVILP
PRD  cccccccccccccchhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhcccccccc

SEQ  KDLKSGQKQKCYFLKFSFGSAQALNLTGKDWKLGKGRHALTPRHLHAWRLGLPPESTR
PRD  cccccccccccccccccccccchhhhhhhhhccccccccccccccccchhhhhhhhhcccccccc

SEQ  LPGLRVVPPPFQEQALQTLKLDHPKIAAWRSRKIGLYNSLCPGTLCILLPGTKSTHG
PRD  cccccccccchhhhhhhhhhhchhhhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ  SLSGLGLMGIKEEESAGPGLCS
PRD  cccccchhhhhhhhhcccccccc

```

Prosites for DKF2phtes3_15j3.2

PS00001	219->223	ASN_GLYCOSYLATION	PDOC00001
PS00001	419->423	ASN_GLYCOSYLATION	PDOC00001
PS00002	723->727	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	182->185	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	279->282	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	447->450	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	605->608	PKC_PHOSPHO_SITE	PDOC00005
PS00005	630->633	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	658->661	PKC_PHOSPHO_SITE	PDOC00005
PS00005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	421->425	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	630->634	CK2_PHOSPHO_SITE	PDOC00006
PS00007	370->379	TYR_PHOSPHO_SITE	PDOC00007
PS00008	27->33	MYRISTYL	PDOC00008
PS00008	186->192	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	714->720	MYRISTYL	PDOC00008
PS00008	720->726	MYRISTYL	PDOC00008
PS00009	337->341	AMIDATION	PDOC00009

Pfam for DKF2phtes3_15j3.2

```

HMM_NAME  RNA recognition motif. (aka RRM, RBD, or RNP domain)
HMM        *IYVGNLPWDTTEEDLRdLfSQFGpIvsIrMMrDRtGRSRGFAFVEFED
Query      571 IYLSGVS-ETFKQLLQEPRLFLGLEAVILPKDLKSGKQKCYFLKFKS 618
HMM        EEDAekAIdemNG...meFmGRrIRV*
Query      619 FGSAQALNLTGKDWKLGKGRHALT 643

```

DKFZphtes3_15k11

group: signal transduction

DKFZphtes3_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
1 GAGCAAGCGG AGCGGCCCTC GCCCAAGCCA AGCGCGCGTG CCAACCTCC
51 CGCCCGCCCG CGCTCTGTCT CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC CGCGGCGGTG CCGGTGGGG
151 TTCTACGACA TCGAGGGGAC GCTGGGCAAG GGCACCTTCG CTCTGGTGAA
201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAATCTA CCGAGAACTA
301 CAAATAATGA AAATGTTAGA CCACCTCAC ATAATCAAA TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATTCTGGCA AATCCTGTCT GCTGTTGATT ATTGTCATGG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAATCTC CTCTGGGATA
551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGGAAA TTCTCTTAAA
601 AGTGGTGAAC TGTGGCAAC ATGGTGTGGC AGCCCCCTTT ATGCAAGCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAGG AGCCAGCTCT GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCTCTGTCT GTGGAGCTCT GCCCTTTGAT
751 GGACCGACTC TTCCAATTTT GAGGCGAGGG GTTCTGGAAG GAAGATTCCG
801 GATTCCGTAT TTCATGTCAG AAGATTGCGA GCACCTTATC CGAAGGATGT
851 TGGTCTTAGA CCCATCCAA CCGCTAACCA TCCTGTCCAG AGCCTCTTTC
901 AAATGGATGC TCATAGAAGT CACTGCGGGA GTTTAATGAG CAGGTCTCTG
951 AGACGAGAA AATGAGCCAT ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1001 GACTGATGCA CAGCCTTGG AATGATGCGA CTTTGTCTGC TTTTGTGGA
1051 CAGAACAAGA GCTATACCA CTTTGTCTGC AGTGGAGCAG AGACTTGATG
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGCACAACCT TGCCAAGGCA
1151 GCGCCAGCGG TCGGCTTAGC ACCATTCTCT GCACCAACCT TGCCAAGGCA
1201 CAGACTGTGG GCTCTCCAGT GACCATGCTG TCACCAACA TGAGGCTGCT
1251 GCGATCTGCC CTCTCCCCC AGCATCCAA CCGGAGGCC TTTTCAATTC
1301 CAGCATCTGG CTGTGAGGCG GAAGCTGCTG TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCTGTGCT CTCCTGTCTT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCAGG CAACATCATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCCTCAT
1501 GCCTTTGAGG CATTTAGTCT CACACCGAGC GGCAGAGAGC GGCACACTCT
1551 GTCAAGAGTG ACCAATCAAC TGTCTGTGAT GCCTGGGCGA GGGAAAATTT
1601 TCTCCATGAA TGACAGCCCC TCCTTGACA GTGTGGACTC TGACTATGAT
1651 ATGGGGTCTG TTTCAGAGGGA CCTGAACCTT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCCATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCCAGCCA TGCAAGGCTT GAGCTCCAG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGCGCCGAG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATTT AGACAACATC
1901 TTCAGAACTC GGCTAGAACC AAAGGAATTC TACCTTAACC TGGGCGCGGC
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GTCCCTTCAG GAAGAGCTTT
2001 GGCTCTCTAG CTCCAGGACC TTGCTAGCAG CTGCCCTCAG GAAGAGCTTT
2051 CTCAGCAGCA GGAAGGCTC TCCACTCTCC CTGCCAGCTT GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCGAGC CTTCTGTCAA AGGCCAGAA CACCTGTGAG CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GGAGACCCAG TACCTGCAGC CTTTATTGCA
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACCTCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCGTTCCTCC CAGGAGAGCT CCACCGCCTT CTCAGCAGGC CCCACCGTTC
2401 AGCCTGACCC AGCCCTTGAG CCCCGTCTGT GAGCCTTCTC CCGAGCAGAT
2451 GCAATACAGC CTTTCTCTCA GCCAGTACCA AGAGATGCAG CTTAGGCCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGTCT CTCCTGTGCC CACGACGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCAACC CCTCCACCA CACGACAGCC
2601 AGGAGCTGCC CCAGCCCTCT TACAGTTCTC CTATCAGACT TGTAGCTGCT
```

```
2651 CAAGCGCTGC TTCCCTGCG CCAGACTATC CCACTCCCTG TCASTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAAGAG
2751 CCCAGGACTG CAAGAGGCCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACCGTGT GGATCCACAA
2851 CACAACGGGT ATGTCTTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTCTCTAT TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGCT GGATGTGTCT TCCTCTGGT
3051 TCTGCCCCAC CACAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTCGAGTGGA GCAAGCTCTC GAGGGCAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATTATCAA GGGCAACCTT GGTGAAAGCA GAAAGGGTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCAGTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCTTGG AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGAGGAGCA GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCTGCACTG TTCTCGTGGA CACTGTGGCC TTGCGGGCAG TCGGTCTCCA
3751 GGGTACCTGT TGTCTCTTTT CGATGTAAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTGGAGC AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAAG ACATGATAAA TGAACTTTTC TGTCCTATGT
3901 GAAGTGGTAG TGGGCTGCCT TTCCCCCAGA TCATGTCTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTATAT TCAATGTCAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTTCCA CCATGTGGGA ATCAGCATTC TTAATTTCTG
4051 TAAAGTTTGG ACTTGTAAAT AAATGTTCAA GTATTACAGC AATATTCAA
4101 GAAAGAACCA CAGATGTGTT AACCATTTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTA ACCAACACTT ACAATTGAGT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTTGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGT
4301 AATTCATAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAACACCTT TTAAGTTGCC TTTCCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGTTCCTATT
4501 CTGGCCATGC GAGCCAGCTC CCTACCAACG TCGGTAACTT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTCGCCA GCGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACCT TGCTCACCGT TGGACTTGGT GTCAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAATAA AAGAACATCA TTTCAATTTAA
4801 AAAAAAAGGG AAAAAAAGGG CGGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTGAATA AAAAAAAG
```

BLAST Results

Entry HSG4921 from database EMBL:
human STS SHGC-37164.
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:
Homo sapiens mRNA for KIAA0781 protein, partial cds.
Score = 10725, P = 0.0e+00, identities = 2145/2145

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959
Category: known protein

```
1 EQAERPSPKP SRAANPPARP RSCPPLAAG PSMVMADGPR HLQRGPEVRVG
51 FYDIEGTLGK GNFAVVKLGR HRITKTEVAI KIIDKSQDLA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVMT KSMILYLVTEY AKNGEIFDYL ANHGRNLNESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPEVF EGQQYEGPOL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSEDCHELI RRMLVLDPK RLITIAQIKH
```



```

301 KMWLIEVPVQ RPVLVPQEQE NEPSIGEFNE QVRLMHSLG IDQOKTIESL
351 QNKSYNHFAA IYFLVLERLK SHRSSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTMH SPNMRLLRSA LLPQASNVEA FSFPASGCQA EAFMEEECV
451 DTPKVNGLCL DPVPPVLVRK GCQSLPSNMM ETSIDEGLT EGEAEDPAH
501 AFEAFQSTRS GRRRHTLSEV TNQLVVMPGA GKIFSMNDSP SLDSVDSEYD
551 MGSVQRDLNF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVHNRSVP SFREGRRASD TSLTQGI VAF RQHQLNLART KGILELNKVPQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLETQ YLQHLRQKPS LLSKAQNTCQ LYCKEPPRSL EQQLQEHRLQ
751 QKRLFLQKQS QLOAYFNQMQ IAESSYPQPS QQLPLPRQET PPPSQAPPFF
801 SLTQPLSPVL EPSSEMQQYS PFLSQYQEMQ LQPLPSTSGP RAAFPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLQFSYQT CELPSAASPA PDYPTPCQYP
901 VDGAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPLGLFD CEMLDVADPQ
951 HNGYVLVN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15kl1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15kl1, frame 1

Report for DKFZphtes3_15kl1.1

```

[LENGTH] 926
[MW] 103915.77
[pI] 5.70
[HOMOL] TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]
 2e-14
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14
 [FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YBR097w] 1e-10
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]
 1e-10
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]
 1e-10
 [FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]
 4e-09
 [FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.
 cerevisiae, YHR079c] 1e-07
 [FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c]
 1e-07
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04
 [BLOCKS] BL00415A Synapsins proteins
 [BLOCKS] BL00239B Receptor tyrosine kinase class II proteins
 [BLOCKS] BL00107A Protein kinases ATP-binding region proteins
 [SCOP] dlgo1_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-78
 [SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 1e-81
 [SCOP] dlkoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 5e-89
 [SCOP] dlkoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 5e-86
 [SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 3e-80
 [SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 6e-70
 [SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 1e-95
 [SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Homo)] 7e-71
 [SCOP] dlydse_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 2e-96
 [SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Homo)] 2e-72
 [SCOP] dlcdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 5e-97
 [SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma] 2e-68
 [SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 3e-53
 [SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 3e-78
 [SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 1e-58
 [EC] 2.7.1.117 Myosin-light-chain kinase 3e-49
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase (NADPH)] kinase 4e-78
 [EC] 2.7.1.138 Phosphorylase kinase 3e-41
 [EC] 2.7.1.137 Protein kinase 7e-45
 [EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42
 [EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78
 [PIRKEW] phosphotransferase 3e-93
 [PIRKEW] nucleus 2e-74
 [PIRKEW] calcium 2e-40
 [PIRKEW] transferase 3e-33
 [PIRKEW] duplication 2e-32
 [PIRKEW] tandem repeat 7e-45
 [PIRKEW] phorbol ester binding 4e-33
 [PIRKEW] zinc 4e-33
 [PIRKEW] ion transport 1e-32
 [PIRKEW] cell cycle control 1e-45
 [PIRKEW] serine/threonine-specific protein kinase 2e-97
 [PIRKEW] oncogene 1e-34
 [PIRKEW] phospholipid binding 2e-32
 [PIRKEW] autophosphorylation 2e-74
 [PIRKEW] brain 6e-36
 [PIRKEW] heterotetramer 8e-38
 [PIRKEW] mitosis 1e-45
 [PIRKEW] polymer 5e-41
 [PIRKEW] magnesium 6e-80
 [PIRKEW] ATP 2e-97
 [PIRKEW] polyprotein 1e-34
 [PIRKEW] alternative initiators 2e-31
 [PIRKEW] phosphoprotein 2e-74
 [PIRKEW] apoptosis 8e-38
 [PIRKEW] cGMP binding 4e-33
 [PIRKEW] glycoprotein 3e-36
 [PIRKEW] skeletal muscle 8e-38
 [PIRKEW] protein kinase 2e-50
 [PIRKEW] testis 5e-41
 [PIRKEW] cAMP binding 8e-38
 [PIRKEW] transforming protein 4e-33
 [PIRKEW] purine nucleotide binding 7e-52
 [PIRKEW] calcium binding 7e-45
 [PIRKEW] alternative splicing 5e-42
 [PIRKEW] P-loop 7e-52
 [PIRKEW] lipoprotein 8e-38
 [PIRKEW] proto-oncogene 4e-33
 [PIRKEW] segmentation 1e-34
 [PIRKEW] core protein 1e-34

```

[PIRKW]      muscle 8e-38
[PIRKW]      myristylation 8e-38
[PIRKW]      EF hand 7e-45
[PIRKW]      cell division 3e-49
[PIRKW]      homodimer 1e-32
[PIRKW]      calmodulin binding 5e-42
[SUPFAM]     ribosomal protein S6 kinase II 1e-34
[SUPFAM]     calcium-dependent protein kinase 7e-45
[SUPFAM]     AMP-activated protein kinase 6e-80
[SUPFAM]     protein kinase akt 3e-36
[SUPFAM]     protein kinase SPK1 7e-41
[SUPFAM]     unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase 5e-42
[SUPFAM]     calmodulin repeat homology 7e-45
[SUPFAM]     cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
[SUPFAM]     protein kinase DUN1 6e-36
[SUPFAM]     protein kinase C zeta 4e-33
[SUPFAM]     Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
[SUPFAM]     death-associated protein kinase 8e-38
[SUPFAM]     pleckstrin repeat homology 3e-36
[SUPFAM]     ankyrin repeat homology 8e-38
[SUPFAM]     protein kinase homology 8e-99
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase II 6e-38
[SUPFAM]     protein kinase C zinc-binding repeat homology 4e-33
[SUPFAM]     protein kinase C delta 2e-32
[SUPFAM]     cGMP-dependent protein kinase 3e-33
[SUPFAM]     protein kinase cdrl 1e-45
[SUPFAM]     kinase-related transforming protein 2e-50
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase I 8e-42
[SUPFAM]     kinase interaction domain homology 7e-41
[SUPFAM]     gag-akt polyprotein 1e-34
[PROSITE]    PROTEIN_KINASE_ATP      1
[PROSITE]    MYRISTYL                3
[PROSITE]    AMIDATION                2
[PROSITE]    CAMP_PHOSPHO_SITE        4
[PROSITE]    CK2_PHOSPHO_SITE         15
[PROSITE]    TYR_PHOSPHO_SITE          2
[PROSITE]    PKC_PHOSPHO_SITE          10
[PROSITE]    ASN_GLYCOSYLATION         2
[PROSITE]    PROTEIN_KINASE_ST        1
[PFAM]       Eukaryotic protein kinase domain
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY      12.31 %

```

```

SEQ      MVMADGPRHLQRPVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDSQSLDAVN
SEG      .....
lctpe    .....EEECTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ      LEKIYREVQIMKMLDHPHIKLYQVMTKSMPLYLVTEYAKNGEIFYDLANHGRNLNESEAR
SEG      .....
lctpe    HHHHHHHHHHHHCCCTTTBCCEEEEETTTTEEEECTTTTBHHHHHHHHHCCCHHHH

SEQ      RKFWQILSAVDYCHGRKIVHRDLKAENLLDNMNIKIADFGFGNFFKSGELLATWCGSP
SEG      .....
lctpe    HHHHHHHHHHHHHHCCCECCCGGGEETTTTCEEECTTTTEETT-TTBC-CCCCCG

SEQ      PYAAPEVFEGQQYEGPQLDIWSMGVVLVLCGALPFDGPTLPILRQRVLEGRFRIPIYFM
SEG      .....
lctpe    GGCHHHHHHCCBC-HHHHHHHHHHHHHHCCCTTTTTHHHHHHHHHHCCCTTTT

SEQ      SEDCEHLIRMLVLDPSKRLTIAQIKHKWMLIEVPVQRPVLPQEENEPSIGEFNEQV
SEG      .....
lctpe    CHHHHHHHHTTTTGGGTTTHHHHHHCGG.....

SEQ      LRLMHSLGIDQOKTIESLQNSYNHFAAIYFLLVERLKSRRSFPVEQRLDGRQRRPSTI
SEG      .....
lctpe    .....

SEQ      AEQTVAKAQTVGLPVTMHSNMLRLRSALLPQASNVEAFSPASGCQAEAAFMEEECVDT
SEG      .....
lctpe    .....

SEQ      PKVNGCLLDPPVPLVRKGCQSLPSNMETSIDEGLETEGEAEEDPAHAFAFAQSTRSGQ
SEG      .....
lctpe    .....XXXXXXXXXX.....

SEQ      RRHTLSEVTNQLVVMFGAGKIFSMNDSPSLDSVDSEYDMGVSQVDLNFLEDNPSLKDIML
SEG      .....
lctpe    .....

```

```

SEQ      ANQPSPRMTSPFISLRPTNPAMQALSSQKREVHNRSPVSEFREGRRASDTSLTQGIVAFRQ
SEG      .....
lctpe    .....

SEQ      HLQNLARTKGILELNKVLLEYQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVST
SEG      .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpe    .....

SEQ      LPASVHPQLSPROSLETQYLQHLRQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQK
SEG      .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpe    .....

SEQ      RLFLQKSQLQAYFNQMQUIAESSYPQPSQQLPLPRQETPPPSQQAPPFSLTQPLSPVLEP
SEG      .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpe    .....

SEQ      SSEQMQYSPFLSQYQEMQLQPLPSTSGPRAAPPLPTQLQQQPPPPPPPPPPRQCAAPA
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
lctpe    .....

SEQ      PLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQQSDLTGPDGPRSPGLQEPSSSYDPLAL
SEG      xxx.....
lctpe    .....

SEQ      SELPGLFDCEMLDAVDPQHNGYVLVN
SEG      .....
lctpe    .....

```

Prosites for DKFZphtes3_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_15k11.1

HMM_NAME Eukaryotic protein kinase domain

HMM		*YeigRilGeGsFGtVYkCiWr.TGeIVAIKIkkrrms.....FlREI	
Query	20	Y I+++G+G+F++V++++R T +VAIKII+K++++ + RE+	68
		YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAVNLEKIYREV	
HMM		qIMRrLnHPNIIRFYDwFedddDHIMIMEYMeGGDLFDYIrrngpMsEw	
Query	69	QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E	117
		QIMKMLDHPHIKLYQVME-TKSMLYLVT EYAKNGEIFDYLANHGRLNES	
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKICDFGLARqM	
Query	118	E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+ ++	167
		EARRKFWQILSAVDYCHGRKIVHRDLKAENLLLDNNMNIKIADFGFGNFF	
HMM		nnYerMttfCGTfWYMMAPeVIimg.nyYttkVDMWSFGCILWEMMTGep	
Query	168	+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G +	215
		KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLOIWSMGVVLYVLVCGAL	
HMM		PFyddnMemImriqrfrfwpnCSeElyDFMtwCwnyDPekRPTFrQI	
Query	216	PF++ ++ + + +++ R++++ +SE++ +++R+++ +DP+RR+T+ QI	265
		PFDGPTLPILRQRVLEGRFRIPYFMSDCEHLIRRLVLDPKRLTIAQI	
HMM		LnHPWF*	
Query	266	+H W+ KEHKWM 271	

DKFZphtes3_17f10

group: testes derived

DKFZphtes3_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1 CTTCAGTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51 TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTTCC AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCC TGAAACATAA ATCGTCGGGA
201 AAGATTTTTC CTAGTGAAAC CCCTGAATTT CAACCAAGCA CAACACAGCA
251 TGAAGAAATT GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
301 CTAAAAAAGG TCCCCAGTA CTTTACGAAG ATGAGCTTAG GGAAGAAGTA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
401 TGAATATAGC CCTCCATCAA CAGAAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGCAGAGA ATTCTGTTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTTCTGCTG AAATTCAGCC TCCATCAGCT
651 GAAGAGTCTC CTTCTGTAGA GCTTCTGGCT GAAATTCCTG CTCCATCAGC
701 TGAAGAGTCC CTTTCAGAA AGCCTCCTCG TGAAATTCCT CCTCCACCAG
751 CTGAAAAATC TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTTA AGTACAGCCT TTACAGCTG AGGGCGCCCT
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCCCTG
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCTAGAGA AGAGGCTCGA
951 GAACTTCAGC TTTCAACAGC TATGGAGACC CCTGCAGAAG AGGCTCCTAC
1001 TGAATTTTCA TCTCCATTAC CTAAGAGAGC CACTGCAGAA GAGGCCTCTG
1051 CTGAAATTCA GCTTCTAGCA GCTACGGAGC CTCTGCAGAA TGAAATCTCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCTC
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTGC AGATGAAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCCTTTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTACAGT TCTAGCAGCT ACAGAGGCTT TGCAGAAGA
1451 GGCTCCTGCT GAAGTTCAGC CTCCACCAGC TGAGGAGGCC CCGCTGAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCTGTA AGTTTCAGCT CCACCACTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGGCCCCGCT GAAGTTCAGC
1651 CTCACCCAGC TGAGGAGGCC CCCCCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCTCG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCTATAGA
1751 AGAGACCCTT GCTGCAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCCT CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTCAC TCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGGT AGAAGATGTG TCTACAGAAT
1951 TTCAGTCACC GCAGGTGGCA GGAATTCAG CAGTAAAAAT AGGATCGGTT
2001 GTTTTGGAAG GTGAAGCAAA ATTTGAAGAG GTTTCAAAAA TCAATCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAAATTAAC AACGTCCTCC TGAAGTGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGTTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGG ACCCAAGAC TTGGAATACA GGTGGAATAA TGAACAATAA
2301 AAAGTGTAGC AGCATAAAAT TACTTGTGTT AATTTTCATC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAATTGCT
2401 TAAAAATACC TTCTACAGT AAAGTGTGTT ACACAGGTAA AGTTTATCT
2451 GCAGCCATCT TTTCTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAG AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710
Category: similarity to known protein
Classification: unclassified

1 MDRSQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP
101 VLLEDELREE VTPVVVQEGS AVKKVASAEI EPPSTEKFPF KIQPLVEEA
151 TAKAEPRPAE ETHVQVQPST EETPDAAEAT AVAENSVKVQ PPPAEAPLV
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPFAEI LPPPAEKSPS
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APKVEPPTV EETLAEVQPL
301 LPEEAPREEA RELQLSTAME TPAEEAPTEF QSPLPKETTA EEASAEIQLL
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAEETT EEASAEIQL
401 LAIEAPADE TPAEAQSPLS EETSAAEAPA EVQSPSAGKV SIEEAPLELQ
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA
501 EEPAEVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE
551 APSEVQPPPA EEPAEVQSL PAEETPIET LAHVHSPPAD DVPAAEASVD
601 KHSPPADLLL TEEFPIGEAS AEVSPPPSEQ TPEDEALVEN VSTEFQSPQV
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGGA PTEIESVVFH
701 IELKQRPPEL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P
= 7.4e-43

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N
= 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat
Length = 1,072

HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43
Identities = 185/622 (29%), Positives = 320/622 (51%)

Query: 33 SESEIVVISRPDSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92
SE +I V+ + + + +E + + + ++ E E Q E G + + TS
Sbjct: 436 SEEKIKVVEKSEKETVIVEEQTEEIQTVEEVTEEDKEAQGEEEEAEEGGEEAATTSPP 495

Query: 93 QETKKGPPVLLEDELREEVTPVVVQEGSAVKKVASAEIEPPSTEKFPKIQPLVEEATA 152
E P + ++EE P + A K + AE + P+ K PA+++ P ++ A
Sbjct: 496 AEEAASPEKETKSPVKEAKSPAEEKSPAEEK-SPAEEKSPAEEKSPAEEKSPA 554

Query: 153 KAEPRPAETHVQVQPSTTEETPDAAEATAVAENSVKVQPPPAEEAP-LVEFPAEIQPPSA 211
+A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+
Sbjct: 555 EAKS-PAE---VK-SPATVKSPAEEKSPAEEKSPAEEKSPAATVKSPGEAKSPAEEKSPA 609

Query: 212 EESP-SVELLAELPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268
+SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+
Sbjct: 610 VKSPVEAKSPAEEKSPASVKSPEAKSPAEEKSPAEEKSPAATVKSPVEAKSPAEEKSPVT 669

Query: 269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLPPEEAPREEARELQLSTAMETPAE-EAP 327
V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P
Sbjct: 670 VKS-PAEA---KSPVEKSPASVKSPEAKSPAGAKSPA-EAKS---PVVAKSPAEEKSP 721

Query: 328 TEFQSPLPKETTAEEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHAEVQS--- 383
E + P ++ AE S A + PA+ ++PAE+SP+ E S E+A + V+
Sbjct: 722 AEAKPPAEAKSPAEEKSP-----AEAKSPAEEKSPAEEKSPV-EVKSPEKAKSPVKEGAK 775

Query: 384 PLAEETTAEEASAEIQLLAIEAPAD-ETPAEAQSPLEET-SAEAPAE-EVQSPSAGKV 440

LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK
 Sbjct: 776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKKEAKSPEKAKTLDVKSPEAKTP 833
 Query: 441 SIEEA--PLELQPPSGEETTA--EASAAIQLLAATEASA---EAPAEVQPPPAEEAPAE 494
 + EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +
 Sbjct: 834 AKKEAKRPADIRSPQVKSPEAKSPEKEETRTKVPKKEEVKSPVEEVKAKEPPKK 893
 Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553
 V+ P EV+ +EAP E Q P AEE + P +++P E + EEA
 Sbjct: 894 VEEETPATPKTEVKESKKDEAPKEAQKPAKEEKEPLTEKP--KDSPEGAKK---EEAKE 948
 Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAEEASVD-KHS 603
 + P EE PA++ ' ++ P AE+ +E + P ++VPA D K
 Sbjct: 949 KKAAPPEETPAKLGVKKEAKPKAEADAKAKEPSKPSKEKEPKKEEVPAAPKPKDTKEE 1008
 Query: 604 PPADLLTTEFFIGEASAEVSP--PSEQT-PEDEALVENSTEFQSPQ 649
 + EE P +A A+ P E + P+ E ++ ST+ + Q
 Sbjct: 1009 KTESKKPEEKPMQAKAKEEDKGLPQEPSPKPKTEKAESSTSDQKDSQ 1057
 Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42
 Identities = 184/628 (29%), Positives = 310/628 (49%)
 Query: 18 IPPVEKVDKEQTYFSESEIVVISRP--DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74
 I VEK +KE ++E + ++ + E+ + + G+ A+ P + A
 Sbjct: 440 IKVVEKSEKETVIVEEQTEEQTEEVTEEDKEAQGESEEEAEGGEEAATTSPPEEA 499
 Query: 75 TNSNEEIQKNISRTSFTQETKKGPPVLEDELREEVTVPVVQEGSAVKKVASAEIEPPS 134
 + +E + + + K P E + E P + A K + AE + P+
 Sbjct: 500 ASPEKET-KSPVKEAKSPAEAKSPA---EAKSPAEEKSPAEEVKSPEAVK-SPAEEKSPA 554
 Query: 135 TEKFPKIQPPLVEEATAKAEPRAEETHVQVQ-PSTEETPDAAATAVAENSVKVQPPP 193
 K PA+++ P ++ A+A+ ++ +V+ P+T ++P + A A++ +V+ P
 Sbjct: 555 EAKSPAEEVKS PATVKS PAEAKSPA EAKSPA EVKSPATVKS PGEAKSPA EAKSPA EVKSPV 614
 Query: 194 AEEAPL-VEFP AEI QPPSAEESPS-VELLAELPPSAEESPS-EPPAEILPPPAEKS 250
 ++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+
 Sbjct: 615 EAKSPA EAKSPASVKS PGEAKSPA EAKSPA EVKSPATVKS PVEAKSPA EVKSPVTVKSPA 674
 Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPEEAPR 307
 + E++SP++ K+P E + P A+ E ++P + P ++ AE + P ++P
 Sbjct: 675 EAKSPVEVKS PASVKS PSEAKSPA GAKSPA EAKSPV VAKSPA EAKSPA EAKSPA EAKSPA 734
 Query: 308 EEARELQLSTAME--TPAE-EAPTEFQSP----LP-KE--TTAEASAEIQLLAATE-- 354
 E + + E +PAE ++P E +SP P KE + AE S E E E
 Sbjct: 735 EAKSPA EAKSPA EAKSPA EVKSP EAKSPA EVKSP EAKSPA EAKSPA EAKSPA EAKSPA 794
 Query: 355 -PPAD-ETPAEARSPLSEET-SAEAAH-EVQSP LAEETTAEAS--AEIQLLAATEAPA 408
 PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ ++P
 Sbjct: 795 KPPAEVKSPEKAKSPNKEEAKSPEKAKTLDVKSPEAKTPEAKRPAADIRSPQVKS 854
 Query: 409 DETPAEQSPLEETSAAE-APA--EVQSPSAKGVSEAPLELQPPSGEETTAEASAA 465
 E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A
 Sbjct: 855 KE---EAKSPEKEETRTKVPKKEEVKSP-----VEEVKAK-EPPKKVE---EETPA 901
 Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525
 E+ +EAP E Q P AEE + P +++P E + A+E A P E
 Sbjct: 902 TPKTEVKESKKDEAPKEAQKPAKEEKEPLTEKP--KDSPEAKKEEAKKAAA---PEE 956
 Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581
 E PA++ + P E+A P++ PSE + P EE PA + +E E+
 Sbjct: 957 ETPAKLGVKKEAKPKAEADAKAKEPSK--PSEKEPKKEEVPAAPKPKDTKEETTESK 1014
 Query: 582 AAVHSPADDVPAEEASVDKHSPPADLL-LTEEFPIGEASAEVSPPPSEQTPEDEA 636
 P EE DK P TE+ ++ + PSE+ PED+A
 Sbjct: 1015 KPEEKPMQAKAKEE---DKGLPQEPSPKPKTEKAESSTSDQKDSQSEKAPEDKA 1067
 Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36
 Identities = 162/540 (30%), Positives = 275/540 (50%)
 Query: 135 TEKFPKIQPPLVEEATAKAEPRAEETHVQVQSTEETPDAAATAVAENSVKV 189
 TE P KI P + K+E + +E+ V V+ TEE E T E +
 Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEEQTEEVTE--EEDKEA 474
 Query: 190 QPPPAEEAPLVEFP AEI QPPSAEESPSVELLAELPPSAE--SPSE-EPPAEILPPPAE 246
 Q EEA A P AEE+ S E E P EE SP+E + PAE P
 Sbjct: 475 QGEEEEAEEGGEEAATTSPPEEAASPE--KETKSPVKEAKSPA EAKSPA EAKSPA EAKSPA 532
 Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
 KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + +P
 Sbjct: 533 KSPA-----EVKSPA EVKSPA EAKSPA EAKSPA EAKSPA EAKSPA EAKSPA EAKSPA 583

Query: 307 REEARELQLSTAME--TFAE-EAPTEFQSPLPKETTAEASAEIQLLAATEPPAD-ETP 361
 E + + E +PAE ++P E +SP+ ++ AE S A ++ + PA+ ++P
 Sbjct: 584 AEVKS PATVKSPGEAKSPAEAKSPAEVKS PVEAKSPAEAKSPASVKS PGEAKSPAEAKSP 643

Query: 362 AEARSPLSEETSAE-EAHAEVQSPLAEETTAEEASAEIQLLAIEAPAD-ETPAEQSPL 419
 AE +SP + ++ E ++ AEV+SP+ ++ AE A + ++ ++PA ++P+EA+SP
 Sbjct: 644 AEVKS PATVKSPVEAKSPAEVKS PVTVKSPAE-AKSPVE----VKSPASVKS PSEAKSP- 697

Query: 420 SEETSAAEAPAEVQSPS-AKGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAE 478
 + ++PAE +SP AK + ++P E +PP+ ++ AE S A A + A A+
 Sbjct: 698 ----AGAKSPAEAKSPVAKSPAEAKSPAEAKPPAEAKSPAEAKSPAE---AKSPAEAK- 749

Query: 479 APAEVQPPPAEEAPAEVQPPPAEEAP--AEVQPPPAEEAPA--EVQPPPAEEAPAEVQPP 534
 +PAE + P ++P + + P E A AE + P ++P E+PP ++P + + P
 Sbjct: 750 SPAEAKSPVEVKSPEKAKSPVKEGAKSLAEAKSPKAKSPVKEEIKPPAEVKSPEKAKSP 809

Query: 535 PAEEAPAEVQPPPAEEAPSEVQPPPAEEA--PAEVQSLPAEETPIETLAHVHSPADDV 592
 EEA + + + E + P EEA PA+++ ++P +E SP ++
 Sbjct: 810 MNEEAKSPKAKTLDVKSPEAKTPAEKAEKRPADIRSPQVKSPEKE---AKSPEKEET 866

Query: 593 PAEEASVDKHS--PPADLLTTEFFIGEASA EVSPPPSEOTPEDEALVENVSTEFOSQOV 650
 E+ + K P + + +E P + + E P + +T E Q P+
 Sbjct: 867 RTEKVAPKKEEVKS PVEEVKAKEPP--KKVEEEKTPATPKTEVKESKKDEAPKEAQKPKA 924

Query: 651 AGIPAVKLGSVVLEGEAKFEEVSK 674
 + GEAK EE +
 Sbjct: 925 EEKEPLTEKPKDSPGEAKKEEAK 948

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34
 Identities = 123/390 (31%), Positives = 213/390 (54%)

Query: 308 EEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPA---EA 364
 E+ E+Q++ E EE E Q +E AEE E A T PPA+E + E
 Sbjct: 455 EQTEEIQVT---EEVTEEDKEAQGE---EEEEAEEGGEEA---ATTSPPAEAAASPEKET 506

Query: 365 RSPLSEETSAAEHAHVQSPLAEETTAEEASAEIQLLAIEAPAD-ETPAEQSPLSEE 422
 +SP+ EE + AE +SP ++ AE S AE+ A ++PA+ ++PAE+SP +
 Sbjct: 507 KSPVKEEAKSP---AEAKSPAEAKSPAEAKSPAEVKS PVEAKSPAEAKSPAEAKSPAEV 563

Query: 423 TSAE-EAPAEVQSPS-AKGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAEAP 480
 + A ++PAE +SP+ AK + ++P ++ P GE + EA + ++ + EA ++P
 Sbjct: 564 SPATVKSPAEAKSPAEAKSPAEVKS PATVKSP-GEAKSPAEAKSPAEVKS PVEA---KSP 619

Query: 481 AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP 540
 AE + P + ++P E + P ++PAEV+ P ++P E + P ++P V+ P ++P
 Sbjct: 620 AEAKSPASVKS PGEAKSPAEAKSPAEVKS PATVKSPVKEAKSPAEVKS PVTVKSPAEAKSP 679

Query: 541 AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIETLAHVHSPAD-DVPAEEASV 599
 EV+ P + ++PSE + P ++PAE +S ++P E A PPA+ PAE S
 Sbjct: 680 VEVKS PAVKS PSEAKSPAGAKSPAEAKSPVAKSPAEAKSPAEAKPPAEAKSPAEAKSP 739

Query: 600 DKHSPPADLLTTEFFIGEASA EVSPPPSEOTPEDEALVENVSTEFOSQOVAGIPAVKLG 659
 + PA+ E ++ EV P ++P E +++ E +SP+ A P VK
 Sbjct: 740 AEAKSPAEAKSPAE---AKSPVEVKSPEKAKSPVKEG-AKSLA-EAKSPAEAKSP-VK-E 792

Query: 660 SVVLEGEAKFEEVSKINSVLKDLSTNDGQAPTEIES 697
 + E K E +K S +K+ + + +A TL+++S
 Sbjct: 793 EIKPPAEVKSPEKAK--SPMKEEAKSPE-KAKTLDVKS 827

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18
 Identities = 124/420 (29%), Positives = 199/420 (47%)

Query: 252 ELLGEIRSPSAQKAPIEVQPLPA-----EGALEEAPAKVEPPTVEETLAEVQPLLPPEAP 306
 ELLG+I+ A +A + + A AL E A++E TV+ TL +
 Sbjct: 236 ELLGQIQCGGAQAQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEEWFRVRLDR 295

Query: 307 REEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPAEARS 366
 EA ++ + AM + EE TE++ L TT E++ L +T+ + +E
 Sbjct: 296 LSEAAKVN-TDAMRSAQEIEI-TEYRRQLQARTT-----ELEALKSTKESLERQRSELED 347

Query: 367 PLSEE-TSAEEAHAHVQSPLAEETTAEEASA--EIQLLAIEAPAD-ETPAEQSPLSEE 422
 + S ++A ++ + L T E A+ E Q L ++ D E A + EE
 Sbjct: 348 RHQVDMASYQDAIQQLDNEL-RNTKWEAAQLREYQDLLNVKMLDIEIAAYRKLEEGEE 406

Query: 423 TSAEEAPAEV-----QSPS-AKGVSI-EAPLELQPPSGEETT-AEASAAIQLLA-A 471
 P+ + PS + + ++ E +++ S +ET EE + IQ+
 Sbjct: 407 CRIGFGSPSFLTEGLPKIPSYSTHIKVKSEKIKVVEKSEKETVIVEEQTEIEIQVTEEV 466

Query: 472 TEASAEAPAEVQPPPAEEAPAEVQPP--PPAEAPAEVQPPPAEEA--PAEVQPPPA 524
 TE +EA E + AEE E PPAEEA + E + P EEA PAE + P
 Sbjct: 467 TEEEDKEAQGE-EEEEAEEGGEEAATTSPPAEAAASPEKETKSPVKEEAKSPAEAKSPAE 525

Query: 525 EEAPAEVQPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAE-ETPIE-ETLA 582
 ++PAE + P ++PAEV+ P ++P+E + P ++PA V+S PAE ++P E ++ A
 Sbjct: 526 AKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 584
 V SP P E S + PA++ E ++ AE P S ++P E ++ E
 Query: 583 AVHSPPADVPAAEEASVDKHSPPADLLLTTEFFIGEASAEVSPPPSEQTP-EDEALVENV 641
 Sbjct: 585 EVKSPATVKSPEGAKSPAEEKSPAEEKSPVE---AKSPAEEKSPAEEKSPAEEKSPA 641
 S E +SP P
 Query: 642 S-TEFQSPQVAGIP 654
 Sbjct: 642 SPAEVKSPATVKSP 655

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18
 Identities = 115/364 (31%), Positives = 166/364 (45%)

Query: 110 EVTVPVQEGSAVKKVASAEIEPPSTEFPAKIQPPLVEEATAKAEPRAE-ETHVQVQ- 167
 E PVV + A K + AE +PP+ K PA+ + P ++ A+A+ PAE ++ V+V+
 Sbjct: 705 EAKSPVAKSPAEEK-SPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 762
 Query: 168 PSTETPDAAATAVAE--NSVKVQPPPAEEA--PL-VEFFPAEIQPPSAEE--SPSVELL 220
 P ++P E A ++AE + K + P EE P V+ P + + P EE SP
 Sbjct: 763 PEKAKSPVKEGAKSLAEAKSPEKAKSPVKEIKPPAEVKSPEKAKSPMKEEAKSPEKAKT 822
 Query: 221 AEILPSPAEEESPSEEP--PAEILPPPAEKSPSVELLGEIRSPSAQKAPIE-VQPLPAE-- 275
 ++ P A+ EE PA+I P KSP+ E E +SP ++ E V P E
 Sbjct: 823 LDVKSPEAKTPAKEEAKRPADIRSEOVKSPAKE---EAKSPEKEETREKVPAPKEEVK 879
 Query: 276 GALEEAPAKVEPPTVEETLAEVQPLPPEEAPREEARELQSTAMETPAEEA-P-TEFQSP 333
 +EE AK P VEE E P P+ +E ++ A + AEE P TE
 Sbjct: 880 SPVEEVKAKEPPKKVEE---EKTPTATPKTEVKESKKDEAPKEAQKPAEKEPLTEKPKD 936
 Query: 334 LPKETTAEEASAEIQLLAATEPPADETPAE--ARSPLEETSAAEAAH-EVQSLAEETT 390
 P E EEA + AA P +ETPA+ + + AE+A A E P +E
 Sbjct: 937 SPGEAKKEEAKK---KAAA--PEEETPAKLGVEEAKPKKAEADAKAKEPSKSEKEKP 991
 Query: 391 A-EEASAEIQLLAIEAPADETPAEAQSPLEETSAAEAPAEVQSPSA-KGVSIIEAPLE 448
 EE A + E E+ + P + + EE Q PS K E++
 Sbjct: 992 KKEEVPAAPKDKTKEKTTESSKKPEEKPMQAKAKEEDKGLPQEPSKPKTEKAEKSST 1051
 Query: 449 LQPPSGEETTAEASAA 465
 Q S A E AA
 Sbjct: 1052 DQKDSQPSEKAPEDKAA 1068

Pedant information for DKFZphtes3_17f10, frame 3

Report for DKFZphtes3_17f10.3

[LENGTH]	710
[MW]	75131.94
[PI]	4.02
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 34.08 %

```

SEQ  MDRSQQTSTRTGYWTMMNIPPVEKVDKEQQTTFSESEIVVISRPDSSSTKSKEDALKHKSS
SEG  .....
PRD  cccccccccccccccccceehhhhhhhccccceeeccccccccchhhhhhhhhccc

SEQ  GKIFASEHPEFQPATNSNEEIGQKNISRTSFTQETKKGPPVLEDELREEVTVPVVQEGS
SEG  .....
PRD  cceeeccccccccccccccccccccccccceeeccccccccchhhhhhhhhheeecccccc

SEQ  AVKVASAEIEPPSTEFPAKIQPPLVEEATAKAEPRAEETHVQVQVQSTETPDAAEAT
SEG  .....
PRD  chhhhhhhccccccccccccccccchhhhhhhhhccccceeeccccccccchhhhh

SEQ  AVAENSVKVQPPPAEEAPLVEFFPAEIQPPSAEESPSVELLAEILPSPAEEESPSEPPAEI
SEG  .....
PRD  hhhhhccccccccccccceeeccccccccccccchhhhhcccccccccccccccccc

SEQ  LPPPAEKSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPL
SEG  .....
PRD  cccccccccccccccccccccccccccccccccchhhhhccccccccccccchhhhh

SEQ  LPEEAPREEARELQSTAMETPAEEAPTEFQSPPLPKETTAEEASAEIQLLAATEPPADET
SEG  .....
PRD  cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccc

```

```
SEQ PAEARSPLSEETSAEEAHAEVQSPLAEETTAEEASAEIQLLAAIEAPAEPTPAEQSPIS
SEG xxxx.....xxxxxxxxxxxxxx.....xxxxxxxxxxxxxx.....xxxx
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc

SEQ EETSAEAPAEVQSPSAKGVSIIEAPLELQPSGTEETAEASAAIQLLAAETASAEAP
SEG xxxxxxxxxx.....xxxxxxxxxx.....xxxxxxxxxx
PRD chhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc

SEQ AEVQPPPAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AEVQPPPAEAPAEVQPPPAEEAPAEVQSPLAETPIETEEAALAVHSPADDVPAEEASVD
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhc ccccccccccccccc

SEQ KHSPPADLLTIEFPFGEASAEVSPPPSEQTPEDEALVENVSTEFQSPQVAGIPAVKLG
SEG cccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ VVLEGEAKFEVYSKINSLVKLDCSTNTDQAPTLIESVFIELKQRPPEL
SEG eeeehhhhhhhhhccccceeeccccccccceeehhhhhhhhhhcccc
PRD eeeehhhhhhhhhccccceeeccccccccceeehhhhhhhhhhcccc
```

(No Prosite data available for DKFZphtes3_17f10.3)

(No Pfam data available for DKFZphtes3_17f10.3)

DKFZphtes3_17117

group: metabolism

DKFZphtes3_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```
1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51 GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGGCAGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTGCTGT CTGTCTCTCT CTTCCACACG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACAACG ACCGGTTCAT CCTCTCCAGG GGACATGCTG
251 CTCTATCCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGAGAA GACACCCCTAC
351 CCCGCGATTG CCGTTTGTGG ACGTGGCAAC AGGGTCCCTA GGTGAGGGAT
401 TAGGTACTGC ATGTGGAATG GCTTATACCT GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCTCT TATGGGAGAT GGCGAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGTCT TTGCTCCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
601 GAGCATGGCG CAGACATCTA CCAGAAATGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGTTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATAC CAAAATCGCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCTTA
951 CAAAGTTGGT GACAAGATAG CTACTCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAACTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAACCTCA CTTTTCTGTA GATATTCAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAAGTGT GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTGATGTA GCTCCGAATG GGAGCCATT TCTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGCCTTGGAG GATCTAGCCA TGTTCGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTGCAACC AGCCAACCAAG
1451 AAAGTGGTCC GCCACGGTGT CAATGATAAA GTACAGTAA TTGGAGCTGG
1501 AAGTTACTCT CATGAAGCCT TAGAAGTGC TGACCATCTT TCTCAACAAG
1551 GTATTCTCTG CCGTGTATC GACCAATTA CCATTAAACC CTTGGATGCC
1601 GCCACCATCA TCTCCAGTGC AAAAGCCACA GCGGCGCGAG TTATCACAGT
1651 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1701 TCTCCAGGGA GCCTGATATC CTTGTTTATC AACTGGCAGT GTCAGGAGTG
1751 CCTCAACGTG GGAATACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1801 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1851 TTATTCTTAA AAAGTCAAGT CTATTGGCTT TGGCCAAAA GCACTGGTAT
1901 CTTTGATTA AATTGATGTT TATTGTCACA AAACCATTTT TTATACCTAT
1951 ACAGTTGTAC TGTTCCTTTT AAAGCAAAGC CATTTAACAT CTTTCTTCAT
2001 TCCTAATTTG GAAATTAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2051 TTAATCTGAG TTATTAATGT GGATTTTAAA ATTGTAAGCA ATAGAATAGG
2101 AAATAAAACA ACTACCTAAT ACAAAATATT CTGATAAGAC TACAATATATC
2151 TGACTGAGCT GGGGATTTAA GTAGAGGTAA CTGTATCTTA AATGAGTATG
2201 ATTTCTTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2251 CCAAGTTTGG AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2301 AGTTTACAG ATAATGTTTG ACTGCAGTTG CCTTGGAAAT TCCTCCAAAG
2351 TTTGCCCTCA TCTCTCCTCT ACAGTTTGGG GGTGATGGTG CAGCAGTGGG
2401 ACATCTCTTG ATGCACCACT CTACTTGCTG TCTGTGAAGT GATGAAAGTA
```

2501 TAACTGGTTC TAGTTTGAC ACTACACACA TAGTTTGTG AAGCTTCAGA
 2551 AATGTTTTT CTTTCTCTG TGGCCAAACC AGTTTGTAA TCTGATTATA
 2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAAGCAT TTAAAAATCA
 2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

96214928:
 Amplification of the transketolase gene in desensitization-resistant
 mutant
 Y1 mouse adrenocortical tumor cells.

99123875:
 Properties and functions of the thiamin diphosphate dependent enzyme
 transketolase.

Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626
 Category: strong similarity to known protein
 Classification: Metabolism
 Prosite motifs: ATP_GTP_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS
 51 VLFHTMKYK QTDPEHPDND RFILSRGHAA PILYAAWVEV GOISESDLN
 101 LRKLHSDLER HPTPLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV
 151 FCLMGDGESE EGSVWEAFAP ASHYNLNLV AVFDVNLGQ SGPALEHGA
 201 DIYQNCCEAF GWNTYLVDPGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR
 251 GIPNIEDAEN WHGKVPVKER ADAIVKLIES QIQTNENLIP KSPVEDSPQI
 301 SITDIKMTSP PAYKVVDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST
 351 FSEIFRKEHP ERFIECIIAE QNMVSVALGC ATRGRTIAFA GAFAAFFTRA
 401 FDQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRISIPNCTVF
 451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR
 501 HGVNDKVTVI GAGVTLHEAL EAADHLSQGG ISVRVIDPFT IKPLDAATII
 551 SSKATGGGRV ITVEDHYREG GIGEAACAAV SREPDILVHQ LAVSGVPQRG
 601 KTELLEDMFG ISTRHITAAV TLTLMK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_17l17, frame 1

SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68)., N = 1,
 Score = 2222, P = 2.5e-230

SWISSPROT:TKT_RAT TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score =
 2202, P = 3.3e-228

TREMBL:RN09256_1 product: "transketolase"; Rattus norvegicus
 Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202,
 P = 3.3e-228

SWISSPROT:TKT_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score =
 2200, P = 5.3e-228

>SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).
 Length = 623

HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230
 Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFFHTM+YK DP +
 6 KPDQQKLQALKDTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYAAMVEVGDISESDDLNLRLKHSDLERHPTPLPFVDVATGSL 126
 P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLDGHPVPKQAFDVTATGSL 125

Query: 127 GQGLGTACGMAYTGKYLKASYRVFCLMGDGESESGSVWEAFASFASHYNLDNLVAVFDVN 186
 GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA Y LDNLVA+FD+N

Sbjct: 126 GQGLGAACGMAYTGKYFDKASYRVYCMGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGGPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246
 RLQGS PAPL+H DIYO CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQGS DPAPLQHQVDIYQKRCEAFGWHTIIVDGHSEELCKAFQA---KHQPTAIIAKT 242

Query: 247 FKGRGIPNIEDAENWHGKVPKERADAIVKLIESIQTNENLIPKSPVEDSPQISITDIK 306
 FKGRGI IED E WHGKP+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEIYSQVSKKILATPPQEDAPSVDIANIR 302

Query: 307 MTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVSGDTMNSTFSEIFRKEHPERFIEC 366
 M +PP+YKVGDKIAT+K YGLALAKL A++R+I L GDT NSTFSE+F+KEHP+RFIEC

Sbjct: 303 MPTPPSYKVGDKIATRKAYGLALAKLGHASDRIIALDGDTKNSTFSELFKEHPDRFIEC 362

Query: 367 IIAEQNMVSVALGCATRGRTIAFAGFAFAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426
 IAEQNMVS+A+GCATR RT+ F FAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSIAVGCATRDRTVFFCSTFAAFFTRAFDQIRMAAISSEINLIGSHCGVSG 422

Query: 427 EDGVSQMALEDLAFMRSPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPETAIVIYT 486
 EDG SQMALEDLAFMRSP TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAFMRSPVPMSTVFYPSDGVATEKAVELAANTKGICFIRTSRPNIAIYS 482

Query: 487 PQENFEIQAKVVRHGVNDKVTIVIGAGVTLHEALEADHLSQQGISVRVIDPFTIRPLDA 546
 E+F++QAKVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIRPLD

Sbjct: 483 NNEDFQVQAKVVLKSKDDQVTIVIGAGVTLHEALAAEESLKKDKISIRVLDPFTIRPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYREGGIGEAACAASREPDIHVQLVAGSVQPRGKTSSELL 606
 I+ SA+AT GR++TVEDHY EGGIGEA V AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYEGGIGEAASAAVGEVPGVTVTRLAVSQVRSRGPFAELL 602

Query: 607 DMFGISTRHIIAAV 620
 MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKFZphtes3_17117, frame 1

Report for DKFZphtes3_17117.1

[LENGTH] 626
 [MW] 67877.52
 [PI] 5.90
 [HOMOL] SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0
 [FUNCAT] m outer membrane and cell wall [M. jannaschii, MJO681] 3e-48
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17
 [FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09
 [FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase]
 2e-05
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate
 dehydrogenase] 2e-05
 [BLOCKS] BL00801F
 [BLOCKS] BL00801E
 [BLOCKS] BL00801D Transketolase proteins
 [BLOCKS] BL00801C Transketolase proteins
 [BLOCKS] BL00801B Transketolase proteins
 [BLOCKS] BL00801A Transketolase proteins
 [SCOP] dtrka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domai 1e-21
 [EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11
 [EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10
 [EC] 2.2.1.1 Transketolase 0.0
 [EC] 2.2.1.3 Formaldehyde transketolase 1e-20
 [PIRKW] transferase 0.0
 [PIRKW] flavoprotein 2e-07
 [PIRKW] Calvin cycle 1e-40
 [PIRKW] heterotetramer 2e-07

{PIRKW} pentose phosphate pathway 0.0
 {PIRKW} magnesium 1e-40
 {PIRKW} thiamine pyrophosphate 0.0
 {PIRKW} oxidoreductase 7e-12
 {PIRKW} fatty acid biosynthesis 4e-10
 {PIRKW} mitochondrion 2e-07
 {PIRKW} peroxisome 1e-20
 {PIRKW} homodimer 1e-40
 {SUPFAM} pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
 {SUPFAM} pyruvate dehydrogenase (lipoamide) beta chain 7e-12
 {SUPFAM} ferredoxin 2[4Fe-4S]-related protein 8e-47
 {SUPFAM} thiamine pyrophosphate-binding domain homology 0.0
 {SUPFAM} pyruvate dehydrogenase (lipoamide) 6e-08
 {SUPFAM} ferredoxin 2[4Fe-4S] homology 8e-47
 {SUPFAM} hypothetical protein C2814 2e-21
 {SUPFAM} transketolase 0.0
 {PROSITE} ATP_GTP_A 1
 {PFAM} Transketolase
 {KW} Alpha_Beta
 {KW} 3D
 {KW} LOW_COMPLEXITY 3.04 %

SEQ MMANDAKPDVKTQVLRDTANRLRIHSIRATCAGSGQLTSCCSAAEVVSVLFHTMKYK
 SEG
 lngsBHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCT

SEQ QTDPEHPDNDRFILSRGHAAPILYAWEVVDISEDLLNLRKLHSDLERHPTPRLPFVD
 SEG
 lngsB TTTTTTTTCEEEETGGGHHHHHHHHHHCTTCHHHHTTTTTTTTTTTTTTTTTTTTC

SEQ VATGSLGQGLGTACGMAYTGKYLKASYRVFCLMGDSSEGSWEAFASFASHYNLDNLV
 SEG
 lngsB CCCCCCTHHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHCHHHHHHHHHHHHHCTTTEE

SEQ AVFDVNRLLGQSGPAPLEHGADIYQNCCEAFGWNTYLVGDHVEALCQAFWQASQVKNKPT
 SEG
 lngsB EEEEECCETTEEEGGCCCCCHHHHH-HHHCCEEEETTTTTHHHHHHHHHHHHHHTTTTCE

SEQ AIVAKTFKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQI
 SEG
 lngsB EEEEECTTTTTTCHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ SITDIKMTSPPAYKVGDKIATQKTYGLAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
 SEG
 lngsB HHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHHHTTTTTTEEEETTHHHHCCTTCEEECCG

SEQ ERFIECIIAEQNMVSVLGCATRGRTIAFAGAFAAFTRAQDLRMGAISOANINLIGSH
 SEG
 lngsB GCEETTTTHHHHHHHHHHHHHHTTTTTEEEEGGGGGGHHHHHHHHHHCTTTEEEEC

SEQ CGVSTGEDGVSQMALEDLAFRSIPNCTVFYPSDAISTEHAITYLAANTKGMCFIRTSQPE
 SEG
 lngsB CCGGGTTTTTTTCHHHHHHHCTTTTEEECCCHHHHHHHHHHTTTTCEEECCCCCB

SEQ TAVIYTPQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEADHLSQOGISVRVIDPFT
 SEG
 lngsB CCTTTTCHHHHHCC-CEEEETTTTTTEEEECCHHHHHHHHHHHHHHHHHCCCEEE...

SEQ IKPLDAATIISSAKATGGRVITVEDHYREGGIGEAVCAVSREPDLVHQLAVSGVPQRG
 SEG
 lngsB

SEQ KTSELDMFGISTRHIIAAVTLTMK
 SEG
 lngsB

Prosite for DKF2phtes3_17117.1

PS00017 595->603 ATP_GTP_A PDOC00017

Pfam for DKF2phtes3_17117.1

HMM_NAME Transketolase
 HMM *vNtIriLaMDAVEKANSQHPGaPMGMAPMAHVLWqrmMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL++++M++++DP P+ ANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYKQTDPEHPD	68
HMM		RDRFVLSNGHaCMLLYsMWHlyGYDmpMWDLKQFRQWHSrTPGHFEIghT +DRF+LS GHa+++LY+ W + G +++++DL+++R+++HS++ +HP ++	
Query	69	NDRFILSRGHAAPILYAAMVEVGD-ISESDLLNLRKLHSDLERHPTPLRP	117
HMM		PGVEVTTGPIGGGIAeNaVWMAIAERNLAATYNRPFGDIfDHYTYCFMGDG ++ +V+TG+LGQG++ +++++Y++++ D++++++C+MGDG	
Query	118	FV-DVATGSLGGGLG-----TACGMAYTGKYLDKASYRVFCIMGDG	157
HMM		CLMEGISWEACSLAGHMqLGNWIAFYDDNrISIDGdTdIWfQEDtYakRF + +EG++WEA ++A+H++L+N++A +D NR+++G++++ + D+Y+ +	
Query	158	ESSEGSVWEAFAFASHYNLDNLVAVFDVNRLGQSGPALEHGADIYQNC	207
HMM		EAYGWHVIEVENDGHdVBeIcaAIEeAKaekDRPTLIICRTVIGYGSPlk EA+GW++ +V DGHdV++C A+ +A +K++PT+I ++T++G+G+PN	
Query	208	EAFGWNTYLV--DGHdVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI	255
HMM		QGTHeWHGAPLGeD* ++ + WHG+P +++	
Query	256	EDAENWHGKVPVKE 269	
HMM		*PqWePnddkIATRKASQqLeaiGPaLPeFwGGSADLTSPNLTrWKGmv P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S++++++ ++	
Query	311	PAYKV-GDKIATQKTYGLALAKLGRANERVILSGDTMNSTFSEIFRKE	358
HMM		WFMPPSISTDCynGNwsGRYIHYGIREHgmAIMNGIALHGnFRPYGGT + + R+I++ I+E++M++++ G+A++G+ +++++ G	
Query	359	H-----PERFIECIIAEQNMVSVALGCATRGR-TIAFAGA	392
HMM		FMMfyDYARPAIRMAALMeIPVIVWTHDSIGLGEDGPTHQPVEHLAHR F++F+++A++++RM A++ +++++H++++ GEDG +++++E+LA+FR	
Query	393	FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSMQALDAMFR	442
HMM		aIPNMsvWRPCDgNETayAWylAvErehTPtLiLSRQNLPLQIErNPrqf +IPN +V++P+D+ T+ A YLA++++ +++++S ++ +++++ P +	
Query	443	SIPNCTVFYPSDAISTEHAiYLAANTKGM-CFIRTSQPETAIVIY-PQEN	490
HMM		ekvaRGGYVLkOmdnePDVILIATGSEMELavaAAKLladEGikaRVVSM +++++++V + + + V++I++G++++A++AA+ L+ +GI +RV+++	
Query	491	FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQGGISVRVIDP	538
HMM		PCTeWFD.....kQDeEYReSVLPdhVPqRVaVEmGvtWCWYKYVGqq +++++D +++++R +++DH++ +++++V ++ +++ +	
Query	539	FTIKPLDAATIISAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDI	587
HMM		GaIfGMNrFGESSGKAPpevLYkhFGFTPENI* + +++ +++ ++ +L+ MFG+ +I	
Query	588	VHQLAVSGVPQR---GKTSELDMFGISTRHI 616	

DKFZphtes3_17n12

group: transcription factors

DKFZphtes3_17n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucine-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```
1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGCTCAT GTAAACAATAG TAGATTGTTT TTTTCTCTAA TATTTCTAGC
101 CAGCCCTTAA GTCAGGTGAT GGAACAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GTGGAGGAGG GGAAGAAGA AAAATGGGAA GAATGCTTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCTGTGTC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTTAACTTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
301 GTGGCTCCC ATCTGCCTCT GCACCCCATTA ATGCACAACA AACCTCACTC
351 TGAGGAGCTA CCAACACTTG CAGTACCAT TCAACAGAT GCTGACTGGG
401 ACAGCGTTCT GTCACTCTAG CAAAGAAATGG AATCAGAGAT GCTGACTGGG
451 TGTTCCTTAT ATTCTCTCCG AATATCTCT ACCTCACCAC ATAAGCCTGA
501 CGAAGGGAGT CGGGACCGTG AGATAATGAC CAGTGTACT TTTGAAACCC
551 CAGAGCGCGC CAAAGGAGT CTTGCCGATG TGGTGACAC ACTGAAACAG
601 AGAAGCTTGT AGAAGATGAC TCGCACTGAA CAAGAGGATT CCTCTGCAT
651 GGAATACTA CTTTCAAAGG ATTGGAAGGA AAAAATGGAA AGACTAAATA
701 CCAGTAACTT TCTTGGACAA ATTAAAGGTA CACTGAGAG CCTGGCAGAA
751 AAAGAACGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTACGGGA
801 CCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA CGCTCACAAA
851 TTGAGAAACA ACGGACGCAA ATGGACCTTG CTCGCCACA GCAAGAACAG
901 ATTGCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCTCCG CTCTATGATC
1001 CAATTTTCC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGGATTCC TCTTCCCCC TGGATAACA TACAACACAG GTGATAACTA
1101 CCCCCTACAG TTCATTCCAT CAACAATGGC AGCTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTACG CTGGAGCAA GATGCCATCA ACTCCACAGC CACCAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCAGAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGCAGC AGAGCCTGTA AAGTCCCAA CGTCTCCAC
1401 CCAGAACCTC TCCCAGCCA GCAAAACAG CCCTGTCAAT CTGCCAAA
1451 AAAGCAGCAT CCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAGTCA ACACCAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAGCCAT TCAGGAGGCG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTTG
1751 AGAATTTGGG GCCCGATTG ACGGGAAAGT CAAATGAAGA TGGAAACTG
1801 GGCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGCG CTAACCTACA GCAGTATTAT TGTGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCACGAG TCTACAGGGA CGCCCGCGGC
1951 CGTGCCAGCA GCGAGCCACA CATTAGCGA CCAATGAATG CATTATCGGT
2001 TTGGGCAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCCGACATGC
2051 ATAACCTCAA CATTAGCAA ATCTTAGGAT CTGCTGGAA ATCAATGTCC
2101 AACAGGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAAACCCGCA CCGAAACGCA
2201 CTGCTATTGT TGATGGCAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGCG TTCTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACA CAGGAACAGG TGTGTGTAT CCTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGCTCT
```

```

2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAAATG TATGATGACT ATGAAGATGA CCCCAAATCA
2551 GACTATAGCA GTGAAATGA AGCCCGGAG GCTGTCAGTG CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCA
2651 ACAAAGAGTT ATTAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

BLAST Results

No BLAST result

Medline entries

95311974:
A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:
The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804
Category: strong similarity to known protein

```

1 MGRMSSKQAT SPFACAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESENKLC SLYSFRNTST
101 SPHKPOEGSR DREIMTSVTF GTPERRKGS LADVVDLAKK KLEEMTRTEQ
151 EDSSCEKELL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMTIQ
201 LISLREQLLA AHDEQKKLAA SQIEKQRQOM DLARQQEQEI ARQQOQLLQ
251 QHKINLLQQQ IQVQGHMPPL MIPFFPHDQR TLAAAAAQQ GLFPPCITY
301 KPGDNYPVQF IPSTMAAAA SGLSPLQLQQ LVAAQLASHQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSFVTQVKD EAAAOPLNLS SRPKTAEPVK
401 SPTSPTQNLF PASKTSFVNL FNKSSIPSP IGGSLGRCSL GKWKSHQHEE
451 TYELDILSSL NSPALFGDQD TVMKAIQEAR KMREQIQREQ QQQPHGVG
501 KLSNNHMLG NSCRNEKERT RFENLGPQLT GKSNEQKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQVYC WPTGGATVAE ARVYRDAGR ASSEPHIKRP
601 MNAFNVWAKD ERRKILQAFP DMHNSNISKI LGRWKSMSN QEQPYVEEQ
651 ARLSKIHLEK YPNYKYPKP KRTCIVDGKK LRIGEYKQLM RSRQEMRQF
701 FTVGQPPQIP ITTGTGVVYP GAITMATTPP SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMEMY DDYEDDPKSD YSSENEAPEA
801 VSAH

```

BLASTP hits

Entry MMSOXL22.1 from database TREMBL:
product: "SOX-L2"; Mouse mRNA for SOX-L2, complete cds.
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:
SOX-L2 - rainbow trout
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:
SOX6 protein - mouse
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330.1 from database TREMBL:
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604.1 from database TREMBL:
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281

Alert BLASTP hits for DKFZphtes3_17n12, frame 1
 No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n12, frame 1

Report for DKFZphtes3_17n12.1

[LENGTH] 804
 [MW] 89332.69
 [PI] 6.97
 [HOMOL] TREMBL:MMSOXLZ2_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0

[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04
 [SCOP] dihmf_1.20.1.1.1 HMG1, fragments A and B [rat/hamster (Rattus)] 1e-13
 [SCOP] dllefa_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mouse] 4e-15
 [SCOP] dlhrya_1.20.1.1.4 SRY [Human (Homo sapiens)] 7e-17
 [PIRKB] DNA binding 4e-94
 [PIRKB] T-cell receptor 4e-07
 [PIRKB] leucine zipper 1e-38
 [PIRKB] alternative splicing 2e-07
 [PIRKB] transcription factor 4e-16
 [PIRKB] transcription regulation 1e-12
 [SUPFAM] HMG box homology 0.0
 [SUPFAM] unassigned HMG box proteins 4e-94
 [PROSITE] ATP_GTP_A 1
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] HMG (high mobility group) box
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 13.81 %
 [KW] COILED_COIL 3.48 %

SEQ MGRMSSQKATSPFACAADGEDAMTQDLTSREKEEGSDQHVASHLPLHPIMHNKPHSEELP
 SEG
 COILS
 lnhm-

SEQ TLVSTIQDADWDVLSQQRMESNNKLCSLYSFRNTSTSPHKPDEGSRDREIMTSVTF
 SEG
 COILS
 lnhm-

SEQ GTPERRKGLADVDTLKQKKLEEMTRTEQEDSSCKEKLKSKDWKEKMERLNTSELLGEI
 SEG
 COILS
 lnhm-

SEQ KGTPESLAEKERQLSTMITQLISLREQLLAHDEQKKLAASQIEKQRQCMDLARQQQEQI
 SEG
 COILS
 lnhm-

SEQ ARQQQQLLQQQHKINLQQQIQVQGHMPLMIPFPHDQRTLAAAAAQGGFLFPPGITY
 SEG
 COILS
 lnhm-

SEQ KPGDNYPVQFIPSTMAAAASGLSPQLQLYAAQLASMQVSPGAKMPSTPQPPNTAGTV
 SEG
 COILS
 lnhm-

```

COILS .....
lnhm- .....

SEQ      SPTGIKNEKRGTSPVTQVKDEAAQPLNLSSRPKTAEPVKSPTSPTQNLFPASKTSPVNL
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PNKSSIPSPIGGSLGRGSSLGKWKSOHQEETVELDILSSLNSPALFGDQDTVMKAIQEAR
SEG      ...XXXXXXXXXXXXXXXXXXXXX.....
COILS    .....
lnhm-    .....

SEQ      KMRREQIQREQQQQPHGVDGKLSSINNMGLNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG      ..XXXXXXXXXXXXXXXXXXXXX.....
COILS    .....
lnhm-    .....

SEQ      PGVIDLTRPEDAEGSKAMNGSAAKLQOYYCWPTGGATVAEARVYRDARGRASSEPHIKRP
SEG      .....
COILS    .....
lnhm-    .....CCC

SEQ      MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEKQPYEEQARLSKIHLEK
SEG      .....
COILS    .....
lnhm-    .....CCCHHHHHHHHHHHHHHHHTTTTCHHHHHHHHHHHHTTTTTHHHHHHHHHHHHHHHHHHH

SEQ      YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRRQEMRQFFTVGQQPQIPITTTGTGVVYP
SEG      XXXXXXXXXXXXXXXX.....
COILS    .....
lnhm-    .....HHHTTTTTTTT.....

SEQ      GAITMATTTPSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGSLAGNEMINGEDEMXY
SEG      .....
COILS    .....
lnhm-    .....

SEQ      DDYEDDPKSDYSSENEAPEAVSAN
SEG      XXXXXXXX.....
COILS    .....
lnhm-    .....

```

Prosite for DKFZphtes3_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphtes3_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWQEMReIKaENPNdmhNtEISKMIgEMWKnMsEEEkM +KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISILGSRWKSMSNQEKQ	644
HMM	PYEdMAeeEKqRYMKEMPeYK*		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKF2phtes3_17n18

group: intracellular transport and trafficking

DKF2phtes3_17n18 encodes a novel 732 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein contains TONB_DEPENDENT_REC_1 Pattern and ATP_GTP_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp

Poly A stretch at pos. 2806, no polyadenylation signal found

```
1 GTCCTTTTAA GTCAGTAAAT TGAACAAAGT CGGTTATTGG GCAAGCAGTT
51 CCTATAAAAA ACTACATGGC TAAGGTTCTT AATGATTGAC CACAAGCAGA
101 TCTTTACCCC TCGGATCTCT AGCTACAAAA GGTCCCCACA CTGAAGAAGC
151 CACTACCTCC ACCACCAACA GCACCAACAC GTCCAGTGCT GCTGGCAACC
201 ACTGGGGCAG CCAAGCGGTC CACCCTCTCT CCCACCATGG CCCGTGAGT
251 GCGCACCCAC CAGGAGACCC TGAACAGGTT TCAGCAGCAG TCCATCCACC
301 TGCTGACGGA GCTCCTCAGA CTGAAGATGA AGGCCATGGT GAGAGTCTAT
351 TCGGTGGGTG CCAAGCCCTT GGACATCACC AGGCCGTTTG TGGAGGCCAG
401 CCAGCTCCTC CACCTCAATG CCAAGGAGAT GGCCTTCAAC TGCTGATCA
451 GCACAGCCGG GAGAAGTGGC TACAGCAGCG GACAGTTGTT GAAAGAGTCC
501 CTCGCAAAAC TGTCCGCCAT TGGGTTGAAC TCGCCCTTAC AGCTGATCTA
551 CCACTCTTCC ACAGCCTGTC TGAGCTTTTC TCTCTCTGCT GGAAGAAGAG
601 CCAAGAAGAA AATAGGCAAA TCTAGAACTA CAGAGATGCT CAGCATCCCG
651 CCCCTGCAAT GAGGAGTGGG AACCCCTGCC AACAGCCTGG ACTTCAGCGA
701 CCCCTGCCCT GAGGCCCGGG AGAGCTGCGA GGAGTTGTGT CCCCACATAG
751 AAGCTGAAAG GGCACATGGG AAAGGGAGGA ATATCTCTTA CCCCATGATC
801 TTACGAAACT ACAAGGCAAA GATGCCCTCT CATCTAATGT TGGCCCGCAA
851 AGGAGACTCT CAGACCCCGG GTTTACATTA CCCTCCACT GCAGGTGCTC
901 AGACTCTCAG CCCACCTCTT CACCCTCTT CTGCCAACA TCATTTGAGT
951 CAGCATTTGT AAGAGGGGAA GGCACCCAAG AAGGCCCTCA AGTTTCATTA
1001 CACCTTCTAT GATGGCTCCT CCTTCGTTTA CTATCCCTCT GGAAACGTCG
1051 CTGTATGTCA GATCCCAACA TGCTGCAGAG GGAGAACCAT CACCTGCTCT
1101 TTTAATGACA TACCTGGATT CTCCTTGCTG GCCCTATTCA ATACTGAAGG
1151 CCAGGGCTGT GTTCACTACA ACCTAAGAAC CAGTTGCCCA TATGTCTTAA
1201 TCTTGGATGA GGAAGGTGGG ACCACCAATG ACCAGCAGGG CTATGTAGTC
1251 CACAAGTGGA GCTGGAATT CAGGACAGAG ACCCTGCTTT CCCTGGAATA
1301 CAAGGTGAAT GAGGAATGA AACTAAAGGT ACTGGGACAG GACTCCATCA
1351 CAGTCACTTT CACCTCCCTG AATGAGACAG TAACACTCAC TGTGTGGGCC
1401 AACAATTGTC CCCATGGAAT GGCATATGAC AAACGGCTGA ACCGAGAAAT
1451 CAGCAACATG GACGACAAGG TGTATAAGAT GAGCCGAGCC CTGGCTGAGA
1501 TCAAGAAGCG GTTTCAGAAG ACAGTGACTC AGTTTATTAA TTCTATCTTG
1551 CTGGCCGCGG GTCTGTTTAC CATTGAATAT CCCACCAAAA AGGAGGAGGA
1601 AGAATTTGTT CGGTTCAGA TGAGATCCAG AACTCATCCC GAGCGGCTCC
1651 CCAAGCTAAG TTTTACTTCA GGAGAAAGTC TTTTACGATC TCAGTCAGGC
1701 CACCTGGAAT CCTCAATTGC AGAGACTTTG AAGGATGAGC CTGAGTCTGC
1751 TCCTGTGAGC CCAGTTCGGA AGACCACCAA AATCCACACC AAAGCCAAGG
1801 TCACATCCAG AGGGAAGGCC CGCGAGGGGC GCAGCCCCAC CAGGTGGGCG
1851 GCCTTGCCCT CAGACTGCCC GCTGGTGCTG CGGAAGCTCA TGCTCAAGGA
1901 AGACACCCGT GCTGGCTGCA AGTGCTGGT GAAGGCCTCC CTGGTCTCTG
1951 ACGTGGAGCT GGAGCGCTTC CTGTTGGCGC CCGAGAGCCC CAGCCAAGTG
2001 CTGGTGTGTT GGATCATCTC AAGCCAGAAC TACACCAGCA CTGGGCAGCT
2051 CCAGTGGCTG CTGAACACTC TCTACAACCA CCAGCAGCGG GGCCGTGGCT
2101 CCCCCTGCAT CCAAGTCCGG TATGACTCCT ACCGCCTGCT GCAGTATGAC
2151 CTGGACAGCC CCCTGCAGGA GGACCTCCCT CTGATGGTGA AGAAGAATCT
2201 TGTGGTGCAG GGGATGATTG TGATGTTTGC CGGGGGGAAG CTCATTTTGT
2251 GGGGCGGTGT TTTGAATGGA TATGGCTCTA GCAAGCAGAA TCTGCTGAAA
2301 CAGATCTTCC GGTCTCAACA GGATTACAAG ATGGGCTACT TCCTGCCGGA
2351 TGACTACAAA TTCAGTGTTC CCAACTTGT CTTGAGCCTG GAGGATCTGT
2401 AATCAGTCAA GAAAGCCGAG TCAGAAGATA TCCAAGGAAG CAGGCTCTCA
```

WO 01/12659

PCT/IB00/01496

```
2451 TTGGCCCTGG AAGACTATGT GGAAGAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG GCCTCAAGA AGTAGGCCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGCCTGGGGT GCTGGGGCTT CTGCCCAGCC
2651 CAGCCCTGCC TCCCCTGGTCT CCCACCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGGC TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2851 CCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782
Category: putative protein
Prosites motifs: ATP_GTP_A (122-130)
TONB_DEPENDENT_REC_1 (1-44)

```
1 MARQVRTHQE TLNRFOQQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSYGS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSTA CLSFSLSAGK EAKKIKGSR TTEDVSMPL HRVGTPPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTAG AQTLSPTSHP SSANHHFSQH COEGKAPKKA
251 FKHYTFYDG SSFVYYPNGN VAVCQIPTCC RGRITITLFP DIPGFSLLAL
301 FNTEGQGCVE YNLKTSCTPV LILDEEGTGT NDQGGYVVHK WSWTSRTETL
351 LSLEYKVNNE MLKLVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRRISNMDO KVKYMSRALA EIKRRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEEFVRF KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTHIHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLRK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPSQVLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCTQCRVD SYRLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSKQNLKQI FRSQDYKMG
701 YFLPDYKFS VPNSVLSLED SESVKAASE DIQSSSSSLA LEDYVERELS
751 LEAEKTRPE VELHPLSRDS KITSKKQAS KR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17n18, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3_17n18, frame 3

Report for DKFZphtes3_17n18.3

```
[LENGTH] 782
[MW] 88030.16
[pI] 9.22
[BLOCKS] BL00286 Squash family of serine protease inhibitors proteins
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 4
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 4
[KW] Alpha_Beta
```


WO 01/12659

PCT/IB00/01496

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKF2phtes3_17n18.3)

DKF2phtes3_18f3

group: testes derived

DKF2phtes3_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```
1 GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51 CCCGACGCGC TGGCGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAACG TGGCCGCCAG CTGCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTAC
251 CCTGGGGACC TCGCTGCTGG TGTGGGCGGT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTGCTGAT CTCTGCAAC
351 TCCCGGGAGC TGGCGAGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGGCGGA GGAACGCCCT CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCTCAT
551 CCCAGGCGCG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCTGACAC CGGGGCTCTG
651 GACGAATCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CAGGACCTCA AGATCTCTGC TGACCAGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAATCATCC
801 TCATGGGATG CTCAGAAATT TGTAAGTCCC TTAGGAAAC ACCAAGCTGG
851 GTTAGGAGCC GAAGGCAAAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCG
901 GTCCCCAAAG CCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCAAAAACCT
1001 CCTTTTCCCT TATCAAAAAC TTTCTGTCTA AACACAGCTG GGCAGGCACT
1051 CCTGTTTAA AGTTATTTTC GGGTCCCTGA CCCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TCGGTGAGGA AGGCATTTCG CTCTATTCCA
1201 GAATTTCTGA TACAAAGAAC TCCAGAAATC AGAGCAATTC AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCCAGA GAGGCAATTC GTGCTGATAT
1301 CTCCTCTCTT TCCTTTTCCC CTCAGCTTAC TTAATCCAGT ATGCGGCCGT
1351 GGTATGAAGT AGGCTTTTCC TGACTGGCTC CCAATCCAGT CCTCAAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTATC TAAGTCTGCG TCCTCACAT
1451 GAAGCTGAGG CCGAGATAGA TGGAGGGACT GCCAATCTCA TTCCCGACA
1501 TCATTGTGTT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCTGGGG
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCTGTGT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CTTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAAAT
1751 CTTGGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTCTCTT GGGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA AATCACTACA ACAAAATGCTA ACTTCTAATC CCCTTGATGA
1901 GCTTTCACGA AGTCTCACGG CTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTCACAGA TGACCAAGGA CAGACTGTGT CCCAGAAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCAGCCGTA CGTGACCTCT GGGGAGTGT CTCACCGTAT
2051 GAATAAGGGA TGTAACTACA AAAGCCCATT AGGGGGCAGT GTTCCCGGCC
2101 TGTTGTAGAA ACTGGTACAG AAAGGATCCT ATATGAAGTT CCTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAAA GTGCCAGTCC ATGTATTTTT
2201 TATTATTTT AAGTTTGTAA TTTAATTTTT AATTATTGTT TAGTGTGTTG
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTCTCTA GTGTTTTAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAGATAA AAGGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAACTTGCA AGCCTGATGT CCTATCAAGT TATGTCTTCT GGGTGACAGA
2701 CAAAATAGCT TGCTTTATGG TGGTATGATG TTGCATTTTC ACTTTGGGGT
```

```

2751 CTGTAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTG
2801 TTAACAACTG TAATGTTGAA AAATAAGTTG AAAAGCTTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCTCTG TTAATTAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCTT TCCAGTGAT GGGGCAGTAT GTCTGAGGAA
2951 GTATAATTTC AGTACTGGGG TCGGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTC TATAAATTGC AATTGGTCTG TATGCTGTTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTG ATCTCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTCCT TTTTGTAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCAGTCAAA CCTCCGACTC
3201 CTGGGTTCAG GCAATTCTTC TGCTCAGCG TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGATTTTTC AGTAGAGACA
3301 GGGTTTCGCC ATGTTGGCCA GGCTGGTCTT GAATCTCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGTGGG GATTACAGGC ATGAGCCACC
3401 TCGCTGGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATGTGTTTGT TTCACAATCA TTTTAAATCA TTTTGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCCTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCAATGCCT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTC TCTCATATT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATT CAGTGGCCCA CTTTATTCTT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAATCTA AGAAATACAT TTGAATAATT
3801 ATAATTAAT GTTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTTCA TTAGATTGAT TCTAGAAACA AATATTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTTGAAA TGGTATGTTG TAAAAGAGAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTCTGGGG CTGCAATCA
4151 TTTGAATTGT TCTGTTTAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATTCA ATAGAACTGC TCCATTAATA AAATAATCCT TAGCAAGCAT
4251 TCGAATCCCTA ACTGCTTTGA TGCATTGCC CTCGGGCACC TGTCATTCC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAGAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTCA ATATTTTAGT GTGAATTGAG ACACAGAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTTCA TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCCTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

BLAST Results

```

Entry HSG27587 from database EMBL:
human STS SHGC-32548.
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:
human STS EST303564.
Score = 1417, P = 8.7e-58, identities = 285/287

```

Medline entries

No Medline entry

Peptide information for frame 2

```

ORF from the beginning to 580 bp; peptide length: 194
Category: questionable ORF
Classification: no clue

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 2

PIR:CGBOIS collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PA1.1 gene: "COL1A1"; Human proalpha 1 (I) chain of type I procollagen mRNA (partial), N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)
Length = 779

HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPQAALPRSQRG 62
G+ G PG + AR PG GPP PA P GA AP G A A P SQ
Sbjct: 230 GDLGAPGFSGARGERGFPGERGVEGPPGAPRGANGAPGNDGAKGDAGAPGAPGSGQAP 289
Query: 63 QLAERNRPRRRHGAQAQPGHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
L G P RGA PG GD +GA G + G VR L + PG A
Sbjct: 290 GL---QGMFGE-RGAAGLPGPKGDRGDAGPKGADGAPGKDG---VRGLTGPIGPPGPAG 341
Query: 123 GAGDRGHL-P-GP-----DARDPELPRVFLPLAGLRGPPAA 156
GD+G P GP D +P P P AG GPP A
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG----GPAPGAPQAALPRSQRG 61
G G PGAA R P AGPP P P G ++G GPA G P + P G
Sbjct: 434 GATGFPGAA-GRVGPFGPSGNAGPPGPPGAPGKESKGRGETGA-GRPGEVGPFGPPG 491
Query: 62 QLAERNRPRRRHGAQAQPGHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 121
A G P G PG PG RG G +RG R L PG +
Sbjct: 492 P--AGEKGAPGAD-GPAGAPCTPGPQGIAGQGVVGLPGQRGE---RGFPGL---PGPS 541
Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160
G +G R P P + GL GPP + RE
Sbjct: 542 GEPGKQGPSGASGERGPPG---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPQAALPRSQRG-R 62
G G PG AR +A PG A G P A PPG + GP PG P A +G R
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATGFPGAAGRVGPPGPS-GNAGP-PGPPGPAGKESKGR 472
Query: 63 QLAERNRPRRRHGAQAQPGHPGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120
GRP G + PG FG GA G + ++ LPG
Sbjct: 473 GETGPAGRP---GEVGPFGPPGAGEKAGADGAPAGPCTPGPQGIAGQGVVGLPGQ 528
Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154
G+RG LPGP + P +G RGPP
Sbjct: 529 R---GERGFGLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAAWARRAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPQAALPRSQ 60
G G PG + PG A+GP P PPG G G A PG P + P +
Sbjct: 29 GPPGAPGPGQFQGPGEPEGASGPMGPRGPPGPKNGDDGAGKPRPGERGPPGPQ 88
Query: 61 G-RQLAERNRPR--RRHGAQAQPGHPGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115
G R L G P + HRG G GD +G G G + R L
Sbjct: 89 GARGLPCTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGEPSPGENGAPGQMGPRGLPGFP 148
Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
GAA G AG+RG +PGP P AG +GPP A
Sbjct: 149 GPKGAAGEPKKAGERG-VPGPPGAVG--PAGKDCEAGAQGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTA----AGPPRPAAP--PGAAPARGGPAP-GAPQAALPR 57
G AG PGA A PG A AGPP PA P PG G P P GA A P
Sbjct: 374 GFAGPPGADGQPGAKGEFGDAGAKGDAGPPGAPGAPGPPGPIGNVGPAGPKGARGSAGPP 433
Query: 58 SQRGQLAERNRPRRRHGAQAQPGHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G A P G PG PG +G G GR V
Sbjct: 434 GATGFPGAAGRVGPPGPSNAGPPGPPGAPGKESKGRGETGPAGRPGEVGP----- 486
Query: 118 PGAAEGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152
PG AG+G PG D A P P +AG RG
Sbjct: 487 PGPPGPAGEKG-APGADGAPAGPTGPP-GGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPG--AAPAR-GGPAP-GAPAQALPRSQR 62
GE G G A + L P G A G P P A P G P G P P G A + +RG
Sbjct: 194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGV 252

Query: 63 QLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEE 122
+ PR GA G GD A G+ G +G R A L PG
Sbjct: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157
GDRG GP D P V L G GPP A
Sbjct: 308 --GDRGDA-GPKGADGAPGKDTV-RGLTGPIGPPGPA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03
Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPAQALPRSQR 61
NG+ GEAG PG R P A G P A PG RG GA A P +G
Sbjct: 67 NGDDGEAGKPRGR-GERGPPGPQARGLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKG 125

Query: 62 RQLAE-RNGRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSL----ADLL 115
+ NG P + G PG PG A G G G V A
Sbjct: 126 EPGPSGENGAPGQ-MGPRGLPGFPKGAAEPKAGERGVPVPGPGAVGPAGKDGEAGQ 184

Query: 116 QLPGAAGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
PG A AG+RG GP A P F L G GPP A
Sbjct: 185 GPPGPAGPAGERGE-QGP-AGSPG----FQGLPGPAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNGNGEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
E GE G PG R L P G GP A PG A RG P P G A A +
Sbjct: 126 EPGPSGENGAPGQMGPR---GLPGFF-GPKGAAGEPKAGERGVPVPGPGAVGPAGKDGEA 181

Query: 61 GRQLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
G Q P RG G PG G+ G G + DL PG
Sbjct: 182 GAQGPFGPAGPAGERGEQGPAGSPG--FQGLP-GPAGPPGEAGKPGEQGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132
+ G+RG PG
Sbjct: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRQLAE 66
GEAG G A R A PG G P P P G A GP PGA Q + + G A+
Sbjct: 347 GEAGPSGPAGTRGA---PGDR-GEPPGPAGFA---GP-PGADGQPGAKGEPGDAGAK 397

Query: 67 RNGRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEGAGD 126
+ P G PG G++ A +GA G G + A + PG + AG
Sbjct: 398 GDAGPPGPAGPAGPPGPIGNVGAPGPKARGSAGPPGATGFFGA-ACRVGPPGPSGNAGP 456

Query: 127 RGHLPDPDARD 137
G PGP ++
Sbjct: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPAQALPRSQRQL 64
G G PGA A G GP P P G A ARG P P Q PR +G
Sbjct: 608 GPPGAPGAPGVPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRGBKGZTG 662

Query: 65 AERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLA-DLLQ-LPG 119
+ + HRG PG PG GA G RG S D L LPG
Sbjct: 663 ZZGBRGIKHGRGFSCLQPPGPGSPGEQGPSGASGPAGPRGPPSAGSPGKDLNGLPG 722

Query: 120 AEGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQ 168
G RG GP A P P P G GPP+ L +P Q
Sbjct: 723 PIGPPGPRGRTGDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02
Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAP---QALPRSQRGR 62
G AG PG A R PG A GP A G A+G P P PA + P G
Sbjct: 152 GAAGEPGKAGERGVPVPGPPG-AVG---AGKDGEAGAQPPGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHGAALQPHGPDLAAGVGRGAGGGHRRRHHHVRSLADLLQLPGAEE 122
Q P G + G PGDL A G G RG R + PG A
Sbjct: 208 QGLPGPAGPPGEAGKPGEGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGFAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154
G G PG D + P G +G P
Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02
Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAARAAALPGT---AAGPPRPAAPPGAAPARG--GPA---PGAPAQALPRSOR 60
G G PG + PG A+GP P PFG G G A PG P + P +
Sbjct: 29 GPPGAPGPGQGPGPEPGEPEGASGPMGPRGPPGPGKNGDDGAGKPGRPGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHGAALQPHGPDLAAGVGRGAGGGHRRRHHHVRSLADLLQL 117
G R L G P + HRG G GD +G G G + L
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAAGDAGPAGKGEPSGPGENGAPQMGPRG-LPGF 147

Query: 118 PGAAEGAGDRG 128
PG AG+ G
Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02
Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAARAAALPGT---AAGPPRPAAPPGAAPARG--GPA---PGAPAQAL 55
G G PGA R A PG A G P P G + RG GPA P PA A
Sbjct: 587 GRDGSFGAKDRGETGFAGAPGPPGAPGAPGVGPAKSGDRGETGPAGPIGPVGPAGAR 646

Query: 56 -----PRSQGRQLAERNRPRRRHGAALQPHGPDLAAGVGRGAGGGHRRRHHHV 108
PR +G + + + HRG G PG + +G G G
Sbjct: 647 GPAGPQGPBGKGTGZZGBRGKGRGFSGLQGPPGPGSGEQGSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAAGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LPGP P PR AG GPP
Sbjct: 706 -----PGSAGSPKDGGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02
Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAARAAALPGTAA----AGPPRPAAPPGAAPARGGAP-CAPAQALPRSOR 60
G G G R AA LPG AGP PG RG P G P A +
Sbjct: 287 GAGLQGMGERGAAGLPGKGDGRDAGPKGADGAPKDVRLGTGPIGPPGAPAGDK 346

Query: 61 GRQLAERNRPRRRHGA---LAQPHGPDLAAGVGRGAGGGHRRRHHHVRSLADLLQL 117
G A +G P RGA +PG PG GA G + D
Sbjct: 347 GE--AGPSG-PAGTRGAPGRGEPGPPGAGFAGPPGADQPGAKGEPDAGAKDAGP- 402

Query: 118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159
PG A AG G + A P+ R G G P AA R
Sbjct: 403 GPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02
Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGAPGAPGAP-QALPRSQGR 62
+G G PGA + PG G PA PG A G P P PA ++ R + G
Sbjct: 574 SREGAPGAEGSPGRDGSFGAKGDRGETGPAGAPGPPGAPGAPGVGPAKSGDRGETGP 633

Query: 63 QLAERNRPRRRHGAALQPHGPDLAAGVGRGAGGGHRRRHHHVRSLADLLQLPGAEE 122
P RG G G+ +G G RG HR + L PG
Sbjct: 634 AGPIGPVGPAGARGPAGPQGRGB-----KGZTGZZGBRGKGRGFSGLQGPPGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
G+G P A P AG RGPP +A
Sbjct: 687 SPGEQG--PS-GASGP-----AGPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02
Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPPRPAAPPGAAPARGGPA-PGAPAQALPRSQGRQLAERNRPRRRH--GALAQ 80
P G P P PG +G P PG P + P RG G P ++ G +
Sbjct: 21 PSGRGLPGPPGAPGPGQGFQGPPEPGEPEGASGPMGPRGPP-----GPPKNGDDGAGK 75

Query: 81 PGHPGDLAA-GV--GRGAGGGHRRRHHHVRSLADLLQLPGAAGAGDRGH--LPGPDA 135
PG PG+ G RG G G HR + L GA AG +G PG +
Sbjct: 76 GRRPGERGPPGPGARGLPGTAGLPGMKGH-RGFSGLDGAAGDAGPAGKGEPSGPGENG 134

Query: 136 RDEL-PRVFLPLAGLRGPPAAA 157
++ PR LP G GP AA

Sbjct: 135 APGQMGPGRG-LP--GFPGPKGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPAQALPRSQRGRLA 65
GEAG G A R A G GPP PA G A G P A G P A + G
Sbjct: 347 GEAGPSGPGAGTRGAPGDRGEP-GPPGPGAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405

Query: 66 ERNGRPRRRHRLAQPCHPGDLAAGVGRGAGGGHRRGR--HHHVRSLADLLQLPGA-- 121
P G + PG G + GA G GR A PG A
Sbjct: 406 AGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPGSGNAGPPGPPGPAGK 465

Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156
EG+ G RG GP R E+ P AG +G P A
Sbjct: 466 EGSKGPRGET-GPAGRPGEVGP GPPGPGAGEKAPGA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARGGPAPGAPAQAL-PRSQR 60
G G PGA R A PG A G P P P G + RG P P + P R
Sbjct: 587 GRDGSFGAKGDRGETGFAGAPGPPGAPGAPGVPAGKSGDRGETGFAGPIGVGPAGAR 646

Query: 61 GRQLAERNRPRRRHRLAQPCHPGDLA-AGVG--RGAGGGHRRGRH--HHVRSLADLL 115
G A G PR +G + G G G +G G G A
Sbjct: 647 GP--AGPQG-PRBGKZTGZGZBGRGKGRGSGLGQPPGPPGSGEQGPGSGASGPAGPR 703

Query: 116 QLPGAAGAGDRG--HLPDPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LPGP P PR AG GPP
Sbjct: 704 GPPGSAGSPGKDGGLNGLPGPIG--PPGPRGRTGDAGAPGP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRLA 65
G G PG A + A G A P P P G A RG G P Q R +RG L
Sbjct: 485 GPPGPPGPGAGEKAPGADGAPGAPGTPG-PQGIAGQRG--VVGLPGQ--RGERGFPLP 538

Query: 66 ERNGRPRRH--RGALAQPCHPGDLA----AGV---GR-GAGGGHRRGRHHHVRSLADL 114
+G P + GA + G PG + AG GR GA G GR + D
Sbjct: 539 GPSGEPGKQGPSGASGERGPPGMPGLAGPPGESREGAPGAEGSPGRDGSFGAKGDR 598

Query: 115 LQL-PGAAGAGDRGHLPGP 133
+ P A G PGP
Sbjct: 599 GETGPAGAPGPPGAPGAPGP 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARG--GPAP--GAPAQALPRSQR 60
G+AG GA A + G GPP PA PG G GPA GAP R +
Sbjct: 311 GDAGPKGADGAPGKDGVRGLTGPIPPGAPGAPGDKGAGPSGPACTRGAPGD---RGE 367

Query: 61 GRQLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHRRGRHHHVRSLADLLQLPGA 120
P G G PGD A G G G + ++ PG
Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGPPGPIGNVG---APGP 423

Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAVRE 160
G G PG RV P AG GPP A +E
Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGPGSGNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRLAE 66
G+AG PGA ++ A L G G A PG RG P A P R L
Sbjct: 275 GDAGAPGAPGSGQAGPLQGMP-GERGAAGLPGKGDGRDAGPKG-ADGAPGKDGVRGLTG 332

Query: 67 RNRGRPRRRHRLAQPCHPGDLAAGVGRGAGGGHRRGRHHHVRSLADLLQLPGAAGAGD 126
G P G PG G+ G G RG A PGA G
Sbjct: 333 PIGPP---GPAGAPGDKGAGPSGPGAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154
+G PG A+ P P AG GPP
Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

ORF from 12 bp to 755 bp; peptide length: 248
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (17-39)
 LEUCINE_ZIPPER (24-46)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 3

TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802.1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.
 Length = 331

HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVLRLREVARRLRLRRSLVANVAGSSLSATGALAAIVGLSLSPVTLTGTSLLVSA 89
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A
 Sbjct: 91 KIQESIEKLRALANGIEEVHRGCTISNVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150
 Query: 90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIATCQDQMR 132
 G+G+ A IT+ + + +S E + AT D+++
 Sbjct: 151 AGVGLGAASAVTGITTSIVENSYTSSAEAE-ASRLTATSIDRLK 193

Pedant information for DKFZphtes3_18f3, frame 2

Report for DKFZphtes3_18f3.2

{LENGTH} 193
 {MW} 19708.24
 {pI} 11.90
 {KW} All_Alpha
 {KW} LOW_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGAARAAALPGTAAGPPRPAAPGAAPARGGPAPGAPAAQALPRSOR
 SEGXXX...
 PRD cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccchhhhh

SEQ GRQLAERNRGRPRRHGALAQPDPGLAAGVGRGAGGGHSRGRHHVRSADLLQLPGA
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
 PRD hhhhhccchhhhhhhhhcccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQCLLHRLMLTW
 SEGXXXXXXXXXXXXX.....XXXXXXXXXXXXX
 PRD cccccccccccccccccccccccccccccccccccccchhhhhhhhhccccchhhhhhhhh

SEQ LPHPOAGGGGHQ
 SEG XXXXXXXXXXXXX
 PRD ccccccccccc

(No Prosite data available for DKFZphtes3_18f3.2)

(No Pfam data available for DKFZphtes3_18f3.2)

Pedant information for DKFZphtes3_18f3, frame 3

[illegible]

PS00029	17->39	LEUCINE_ZIPPER	PDOC00029
PS00029	24->46	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_18f3.3)

DKFZphtes3_1817

group: cell structure and motility

DKFZphtes3_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```
1 GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCAGCAG CGGACGAGGG
51 AGGTGCCGCC GTGCCCCAGG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
101 TAAGAGATTT GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
151 AGACCTCCCG AAAAATCCCT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
201 ACTTGTCCAG CAAAGTGGCC CAAATCCATG GCATTGTCTT AGTACCCCTG
251 AAAGGAAGCC TCTCGAGCAG CATCCAGTCT ACTTGTCACT TTGAGTCCTA
301 CATTTTGATA CCTGTGGAAG AGCATTTCAC GACCTTAAAT GGAAAGGATG
351 TCTTTATTCA AGGGAACAGG ATTAATTAAG GAGCTGGTTT TGCCCTGTCT
401 CTCTCAGTGC CCATTCTCTT TGAAGAACT TTCTACAATG AAAAAGAAGA
451 GAGTTTCAGC ATCCTGTGTA TAGCCCATCC TTGGAAAAG AGAGAGAGTT
501 CAGAAGAGCC TTGGCACCCC TCAGATCCCT TTCCCTGAA AACCATTGAA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTGG ACAGGAACAT
601 CGCCTCTTTC CATCGAACAT TCCGAGAATG CGAGAGAAAG AGCCTCCGTC
651 ACCACATAGA CTCAGCGAAT GCTCTCTACA CCAATGCCT CCAGCAGCTT
701 CTGAGGGAAT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
751 CCTGATGAAG CAGGCACTGG AGATATACGT CCATCATGAA ATTTACAACC
801 TGATCTTTAA ATACGTGGGG ACCATGGAGG CAAGTGAGGA TGCGGCCCTT
851 AACAATAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAAGA GAGCTGGCTC
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTCGCAAAA
1001 GTGGTGACGC TCATTACACA GTCTCAAGC CAGAGAGTGA ACCTGGAGAC
1051 CATGTGTGCT GATGATCTGC TATCAGTCTT GTTATACCTG CTGTGAAAAA
1101 CGGAGATCCC TAATTGGATG GCAAAATTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGGCAAAGGA TGAAGTGGGA TACTGCTGCT CCTCATTCGA
1201 AGCTGCCATT GAATATATTG GGCAAGGAAG CCTCTCTGCT AAACCCCTG
1251 AGCTGAGGGG ATTTGGAGAC AGGCTGTGCC TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCACCCGAC TGCCGTGTTA AGCACATTGC
1351 ATCAGGTAA CAGAAAGAAG TGGAGAACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTCAAC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTTGAAT GATCCCTCAG TTGTCACTCC
1501 ATTTCCAGA GACGACAGGG GGCACACCCC TCTCCATGTG GCTGCTGTCT
1551 GTGGGCGAGG ATCCCTCATC GACCTCCTGG TTCCAAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCCACCTGG CCTGTAGAA
1651 GGGTACCAG AGCGTGACGC TGCTGTGCT GCACCTAAG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACTGGC CTGCACCTAC
1751 GGCCAGGAG ACTGTGTGAA GGCTCTGGTT TACTACGACG TGGAGTCGTG
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGTGCGA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCTCTA AGTGTGCATT
1951 AAACCTAAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCTGTGTC AGTCCCGGCA GCGCTCCGTG
2051 GACTCCATCA GCCAAGATGC CTCCACTTCC AGCTTCTCTT CCATGTCAGC
2101 CGGCTCAAGG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGCTGAT GGAGATCTAG AAATGGTGCG TTACCTGTTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGCAGC
2251 AGACCCCGAA TTCTGTCAAC CGTTGTGCCA GTGCCCAAG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCTCT CCCGCTGCAT GTGCCGCCCC TGACGGCCCG
2401 GCGGACCTC ATCCGCTCTC TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCTGCCA GCAGGGCCAC
2501 TTTCAAGTGG TGAAGTGTCT GTTAGATTCT AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAACACAGC CCCTCATTTA CGCCTGCTCC GGTGGCCATC
2601 ACGAGCTTGT GGCACTGCTG CTACAGCAGC GGGCTTCCAT TAACGCTTCT
2651 AACAATAAGG GCAACACAGC GCTGCACGAG GCTGTGATTG AAAAGCACGT
```

2701 CTTCTGGTGA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTAC TGTTAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGT TCTTTGCATG
3051 AGCCAGGGAG GCAAGTGTG ACACGTGAGC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACGGATGCTG CGGAGACACA
3201 CGGTAGAGGA TCGGTCGTG TCCAGGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTCCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGTAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTTCAGCAC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCCT CTTCAAAGCT AATGAATACA
3501 ATTGAAACAG ACAAAAATTC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAC CACAGCTAGA ATTCCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAATGTCA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCATT TTGATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTTCTAGAA ACATCCCCTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTGAT TCAAATTCRA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTC CTTTTATAA ATTTCTATAG ATTATACGT TATTTTATG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGTCTGAG TCCGTTTTCA
4251 AATGACCTTG TGATAGGGAA ATGGTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGATTAT TTTCTACAA AGTAATTTAT
4351 GATTCTAAT TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAATAA AAAAAAATAA
4451 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4501 G

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050
Category: similarity to known protein
Classification: Cell structure/motility
Prosites motifs: ATP_GTP_A (945-953)

1 MALYDEDLLK NPFYLAQKC RPDLCSEKVAQ IHGIVLVPCX GSLSSSIQST
51 CQFESYILIP VEEHPTQLNG KDVFIQGNRI KLGAGFACLL SVPILEETTF
101 YNEKEESFSI LCIAHPLEKR ESSEELPLAS DPFSLKTIED VREFLGRHSE
151 RFDNRNIAFHF RTFRECEKRS LRHHIDSANA LYTRCLQQLL RDSHLKMLAK
201 QEAQNLNMQQ AVEIYVHHEI YNLIFKYVGT MEASEDAFNF KITSRLQDLQ
251 QKDIGVKEPF SFNIPRAKRE LAQLNKCTSP QOKLVCLRKV VOLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWMH NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIRQGSLSAK PPESEGFQDR LFLQRMSLL SQMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFDCCD EKLVSGRRLND
451 PSVVTFFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYQS VTLLLLHYKA SAEVQDNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLDI GNEKGDPLH IAARWGYQGV IETLLQNGAS TEIQNRKLET
601 PLKCALNSKI LSVMEAYHLS FERRQKSSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSADPEF CHPLCQCPKC APAQKRLAKV PASGLGVNVT SQDGSPLPHV
751 AALHGRADLI RLLKKGANA GARNADQAVP LHLACQGHF QVVKCLLDSN
801 AKPNKKDLSE NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFVVE LLLHAGASVO VLNRKORTAV DCAEQNSKIM ELLQVVPSCV

901 ASLDDVAETD RKEYVTVKIR KWNNSKLYDL PDEPFTROFY FVHSAGQFKG
951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRNNLPAQS GSHAAEKGNS
1001 DWPERPLQTQ TGPCHRRMLR RHTVEDAVVS QGPEAAGPLS TPQEVSAERS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_1817, frame 2

TREMBL:HSU43965.1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin
G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE
ANKYRIN), N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =
380, P = 8.2e-31

>TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
ankyrin (variant 2.1)
Length = 1,719

HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31
Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
+G+T LH+AA+ GQ ++ LV+ GA VNA G TPL+A Q+ + V LL A+
Sbjct: 77 KGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFLLENGAN 136

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYDVES-CRL----- 558
V +G TPL+A GHE+ V L+ Y + RL
Sbjct: 137 QNVATEDGFTPLAVALQQGHENVVAHLINYGTKGKVRPLALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGTPLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615
D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+
Sbjct: 197 PNPVLSKTGFTPLHIAAHYENLNVAQLLLNRGASVNFTPONGITPLHIA--SRGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673
L +R + E + + ++ S + G+ Q +TK +
Sbjct: 255 V-RLLDGRGAQI-ETTKDELTPHCAARNGHVRRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCPLCQCPKCAPAQRRLAKVPA 732
A GD L+ VR LL++ E ++D T+ P H C R+AKV
Sbjct: 312 ---AAQGDHLDVRLLLQYDAE-IDDI--TLDHLP--LHVAHC-----GHHRVAKVLL 358

Query: 733 S-GLGVNVTSDQGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQ 791
G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH
Sbjct: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHL 418

Query: 792 VVKCLLDSNAKPNKKOLSGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAV 851
+VK LL A PN +++ TPL A GH E+ LLQ+ A +NA T LH A
Sbjct: 419 IVNLLQRGASPNVSNVKVET-LHMAARAGHTEVAKYLLQNKAKVNAKAKDDQDPLHCAA 478

Query: 852 IEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVV 896
H +V+LL + A+ + T + A + + +L ++
Sbjct: 479 RIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30
Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAAVCCQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEV 524
TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +
Sbjct: 274 TPLHCAARNGHVRRISEILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIOD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYDVE-----SCR----- 557
+ TPLH+A GH K L+ + +C+
Sbjct: 334 ITLDHLPPLHVAACHGHHRVAKVLLDQKAPNSRALNGFTPLHIACKKNHVRVMEILLKT 393

Query: 558 ---LDIGNEKGTPLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSV 614
+D E G TPLH+A+ G+ +++ LLQ GAS + N ETPL A + V
Sbjct: 394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLL 674
 + Y L + + + Q+P I + +A T L
 Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGH---TPLH 508

Query: 675 RAVADGDLEMYRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAKRLAKVPASG 734
 A +G +E V LLE ++ A T P H + K A+ L +
 Sbjct: 509 IAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAAYKGVRAELLER----D 559

Query: 735 LGVNVTSQDGSPLHVAALHGRADLIRLLKKHGANAGARNADOAVPLHLACQGHFQVVK 794
 N ++G +PLHVA H D++LLL G + + PLH+A +Q +V +
 Sbjct: 560 AHPNAAAGKGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPAWNNGYTPHIAAKQNQVEVAR 619

Query: 795 CLLDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNALHEAVIEK 854
 LL N + + G TPL A GH E+VALL A+ N N G T LH E
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNKNKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 HV V ++L+ HG V + T + A N K+++ L
 Sbjct: 680 HVPVADVLIKGVMDATTRMGYTPHVASHYGNIKLVKF 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPL-CFDDCEKLVSGRLNDPSVVTFFSRD 460
 HIAS GN V LL + + + PL C + +S L D ++
 Sbjct: 244 HIASRRGNVIMVRLLDRAQIETKTKDELTPHCAARNGHVRISILLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYSVTLLLLHYKA 520
 G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A
 Sbjct: 303 KNGLSPIHMAAQGDHLDVRLLLQYDAEIDDITLOHLTPLHVAACHGHHRVAVLLDKGA 362

Query: 521 SAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGV 580
 + NG TPLH+AC H ++ L+ +D E G TPLH+A+ G+ +
 Sbjct: 363 KPNSRALNGFTPLHIACKKNHVRVMELLK---TGASIDAVTESGLTPLHVASFHGLPI 419

Query: 581 IETLLQNGASTEIQNRKLPKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637
 ++ LLQ GAS + N ETP L A ++++ + + K + P+ R
 Sbjct: 420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ----SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDLEMYRYLLEWTE 693
 ++ + E++ + +AG VE +L + + +T
 Sbjct: 480 IGHNTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKASQACMTKKGFTP 539

Query: 694 EDLEDAEDTVSAAD---PEFCHPLCQ-----CP-KCAPAKRLAKVPA---SGLGVNVTS 741
 + V A+ HP P A L V G + +
 Sbjct: 540 LHVAAYKGVRAELLERDAHPNAAAGKGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPA 599

Query: 742 QDGSPLHVAALHGRADLIRLLKKHGANAGARNADOAVPLHLACQGHFQVVKCLLDSNA 801
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 600 WNGYTPHIAAKQNQVEVARSLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNALHEAVIEKHVFVEL 861
 N + SG TEL GH + +L++HG ++A+ G T LH A ++ +V+
 Sbjct: 660 NGNLGNKSGLTPLHLVAQEGHVPVADVLIKGVMDATTRMGYTPHVASHYGNIKLVKF 719

Query: 862 LLLHGASVQVLNK 874
 LL H A V K
 Sbjct: 720 LLHQADVNATK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQK---MCHPLCFDDCEKLVSGRLNDPSVVTFFS 458
 H+AS G+ K V LL +E + T +K H +++V +N + V +
 Sbjct: 50 HLASKEGHVMVVELLHKEIILETTTKKGNALHIAALAGQ-DEVVRELNVYGANVN--A 106

Query: 459 RDDRGTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYSVTLLLLHY 518
 + +G TPL+AA + + L+ GA N G TPL +A Q+G+++V L++Y
 Sbjct: 107 QSQKGTPLMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQGHENVAHLIN 166

Query: 519 KASAEVQDNNGNTP-LHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGY 577
 +V+ P LH+A ++D A V + D+ ++ G TPLHIA +
 Sbjct: 167 GTKGVKVR-----LPALHIAAR--NDDRTAAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVETLLQNGASTEIQNRKLPKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQS 634
 V + LL GAS + TPL A N ++ ++ E + K P+
 Sbjct: 219 LNVAQLLLNRGASVNFTPQNGITPLHIAARRGNVIMVRLLDRAQIETKTKDELTPHLC 278

Query: 635 PQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMYRYLLEWTE 693
 R+ E + + A +TK + A GD L+ VR LL++
 Sbjct: 279 AARNGHVRISILLDHGAPIQA-----KTKNGLSPIHM---AAQGDHLDVRLLLQYDA 329

Query: 694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729
 E ++D D ++ C H + + P C R + +
 Sbjct: 330 E-IDDITLDHLTPLHVAHCGHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQG 788
 + +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G
 Sbjct: 389 LLLKTGASIDAVTESGLTFLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAG 448

Query: 789 HFQVVKCLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALH 848
 H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH
 Sbjct: 449 HTEVAKYLLQNKAKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLH 508

Query: 849 EAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIM--ELL 893
 A E HV V LL AS + K+ T + A + K+ ELL
 Sbjct: 509 IAAAREGHVETVLALLEKEASQACMTKKGFTPLHVAARYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFFSRDD 461
 H+A+ G + E LL ++ H + PL L +L P +P S
 Sbjct: 541 HVAARYGKVRVAELLERDAH?NAAGKNGLTPLHVAVHNNLDIVKLLPRGGSHPSPAW 600

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
 G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+
 Sbjct: 601 NGYTPHIAAKQNVQEVARSLLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLLSKQAN 660

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGYQGI 581
 + + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++
 Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKGV---MVDATTRMGYTPHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRLKETPL 602
 + LLQ+ A + +L +PL
 Sbjct: 718 KFLQHQADVNAKTKLGYSP 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVV 793
 G VN T Q+G +PLH+A+ G ++RLLL GA + +D+ PLH A + GH ++
 Sbjct: 229 GASVNFPTQNGITPLHIASRAGNVMVRLLDLQGAQIETFKTKDELTPHCAARNHVRIS 288

Query: 794 KCLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALHEAVIE 853
 + LLD A K +G +P+ A G H + V LLLQ+ A I+ T LH A
 Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDITLDHLTPLHVAHAC 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909
 H V ++LL GA + + LN + C + + +MELL AS+D V E+
 Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEMLLKTG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14
 Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFFSRDD 461
 HIA+ G+ + V LL +E +K PL K+ L P +
 Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAARYGKVRVAELLERDAH?NAAGK 567

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
 G TPLHVA ++ LL+ +G ++ ++G TPLH+A ++ V LL Y S
 Sbjct: 568 NGLTPLHVAVHNNLDIVKLLPRGGSHPSPAWNNGYTPHIAAKQNVQEVARSLLQYGG 627

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGYQGI 581
 A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V
 Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLLSKQANG---NLGNKSGLTPLHLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPV-QSPQR 637
 + L++G + R+ TPL A N K++ + + + K +P+ Q+ Q+
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPHVASHYGNIKLVKFLQHQADVNAKTKLGYSPHLQAAQ 744

Query: 638 S-VDSISQ--ESSTSFSSMSAGSRQEETK--DYREVEKLLRAVAD 679
 D ++ ++ S S G+ K Y V +L+ V D
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYSVTDVLKVVTD 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVV 793
 G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V
 Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLLSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684

Query: 794 KCLDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIE 853
 L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNAKTLGYSPLHQAAQQ 744

Query: 854 KHVFVVELLLHHGASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896
 H +V LLL +GAS ++ T + A++ + ++L+VV
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYISVTDVLKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHDKDTVQKMCH--PLCFDDC-EKLVSGRLNDPSVVTFFSR 459
 H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR
 Sbjct: 310 HMAAQGDHLDVRLLLQYDAEIDDIT-LDHLTPLHVAHCGHHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
 G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL
 Sbjct: 368 ALNGFTPLHIACKNHVRVMELELLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query: 520 ASAEVQDNNNGNTPLHIACTYGHEDCVKALVYDVESCRDLIGNEKGDTPHIAARWGYQG 579
 AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+
 Sbjct: 428 ASPNVSNVKTETPLHMAARAGHTVAKYLLQ---NKAKVNAKAKDDOTPLHCAARIGHTN 484

Query: 580 VIETLLQNGASTEIQNRKLTPLKCA 605
 +++ LL+N A+ + TPL A
 Sbjct: 485 MVKLLLENNANPNLATTAGHTPLHIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33
 Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVVKLLDSNAK 802
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 601 NYGTPHIAAKQNVQEVARSLLQYGSANAEVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 803 PNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHVFVVELL 862
 N + SG TPL GH + +L+HG ++A+ G T LH A ++ +V+ L
 Sbjct: 661 GNLGNKSLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 L H A V K + + A Q ++ I+ LL
 Sbjct: 721 LQHQADVNAKTLGYSPLHQAAQGGHTDIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11
 Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVVKCL 796
 + T++ G++ LH+AAL G+ ++R L+ +GAN A++ PL++A Q+ H +VVK L
 Sbjct: 71 LETTTKGNLTALHIAALAGQDEVVRELNYGANVNAQSQKGTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHV 856
 L++ A N G TPL A GH +VA L+ +G ALH A
 Sbjct: 131 LENGANQNVATEDGFTPLAVALQGGHENVAHLINYGTK---GKVRPLALHIAARNDDT 186

Query: 857 FVVELLLHHGASVQVLNKRQRTAVDCAE--QNSKIMELL 893
 +LL + + VL+K T + A +N + +LL
 Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29
 Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASA 522
 GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A
 Sbjct: 503 GHTPLHIAAREGHVETVIALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHP 562

Query: 523 EVQDNNNGNTPLHLACTYGHEDCVKALVYDVESCRDLIGNEKGDTPHIAARWGYQGVIE 582
 NG TPLH+A + + D VK L+ S N G TPLHIAA+ V
 Sbjct: 563 NAAGKNGLTPLHVAVHHNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQNVQEVAR 619

Query: 583 TLLQNGASTEIQNRKLTPLKCA 605
 +LLQ G S ++ TPL A
 Sbjct: 620 SLLQYGSANAEVQGVTPHLA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28
 Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVVKCLL 797
 N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL
 Sbjct: 662 NLGNKSLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHVF 857
 A N K G +PL A GH ++V LLL+GAS N ++ G T L A ++

Sbjct: 722 QHQADVNAKTKLGYSPFHQAQQGHTDIVTLLKNGASPNVSSDGTTPLAIAKRLGYIS 781

Query: 858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917
V ++L + V ++ V + S P V + DV+E + +E ++

Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827

Query: 918 KIRKK 922
K ++

Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCV 545
G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V

Sbjct: 35 GVDINTCNQNGNLHLASKEGHVKMVELLHKEIILETTTKGNTALHIAALAGQDEVV 94

Query: 546 KALVYYDVESCRLDIGNEKGTPLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCA 605
+ LV Y ++ ++KG TPL++AA+ + V++ LL+NGA+ + TPL A

Sbjct: 95 RELVNY---GANVNAQSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVA 151

Query: 606 L 606
L

Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06
Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVPFSDRRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQK 507
L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G

Sbjct: 457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVLLLENNANPNLATTAGHTPLHIAAREG 514

Query: 508 YQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCVKALVYYD----- 552
+ L LL +AS G TPLH+A YG + L+ D

Sbjct: 515 HVETVLALLEKEASQACMTKKGFTPLHVAKYGKVRVAELLERDAHPNAAGKGLTPLH 574

Query: 553 --VESCRLDI-----GNE-----KGTPLHIAARWGYQGVETLLQNGASTEIQNRL 597
V LDI G+ G TPLHIAA+ V +LLQ G S ++

Sbjct: 575 VAVHHNNLDIVKLLPRGSGSPHAWNGYTPLHIAAKQNQVEVARSLLYGGSANAESVQ 634

Query: 598 KETPLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMS-SA 656
TPL A M A LS +Q + +S + ++QE +

Sbjct: 635 GVTPLHLAAQEGHAE-MVALLLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIK 690

Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTSAADPEFCHPLCQ 716
G + T + L A G++++V++LL+ + D+ +A+ + + PL Q

Sbjct: 691 GVMVDATTR--MGYTPLHVASHYGNIKLVKFLQH-QADV-NAKTKLGS-----PLHQ 740

Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSQDGSPLHVA 751
+ + + +G N S DG++PL +A

Sbjct: 741 AAQQGHTDI-VTLLKNGASPNVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07
Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSPLHVAALHGRADLIRLLKHGANAGARNADQAVPLHLACQQGHFQVVKCL 796
V D ++ AA G D L++G + N + LHLA ++GH ++V L

Sbjct: 5 VGFREDAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGNLHLASKEGHVKMVVEL 64

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHELVALLQH GASINASNNKGNTALHEAVIEKHV 856
L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+

Sbjct: 65 LHKEIILETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHL 124

Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885
VV+ LL +GA+ V + T + A Q

Sbjct: 125 EVVKFLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26
Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
+ G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y

Sbjct: 42 NQNGNLHLASKEGHVKMVELLHKEIILETTTKGNTALHIAALAGQDEVVRELNVYG 101

Query: 520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGTPLHIAARWGYQ 579
A+ Q G TPL++A H + VK L+ ++ E G TPL +A + G++

Sbjct: 102 ANVNAQSQKGFPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQOQHEN 158

Query: 580 VIETLLQNGASTEIQ 594
V+ L+ G ++

Sbjct: 159 VVAHLINYGTKGKVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21
 Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTLTLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYVDVE 554
 AT A + G ++ L H + ++ + NG LHLA GH V L++ ++
 Sbjct: 13 ATSFRLAARSG--NLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPHLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSV 614
 L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+
 Sbjct: 71 ---LETTTKKGNTALHIALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEV 127

Query: 615 E 615
 +
 Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01
 Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQGHFQVVKLLDSNAKPNKKDLSGNTPLIYACSGGHHELVAL 828
 + G R AD A A + G+ L + N + +G L A GH ++V
 Sbjct: 4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVVE 63

Query: 829 LLQHGASINASNKNGNTALHEAVIEKHVVFVVELLLHGASVQVLNKRQRTAVDCAEQNSK 888
 LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +
 Sbjct: 64 LLHKEIILETTTKKGNTALHIALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 123

Query: 889 I 889
 +
 Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSROEETKKDYREVEKLLRAV 677
 +RRQ+ E VQ + + + Q + + Q ++ +K++R V
 Sbjct: 1614 DRRQQGQEEQVQEAKNFTFQVVGNEFQNPGEQVTEEQFTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
 Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDSGNTPLIYA 817
 +D+G T L+YA
 Sbjct: 1186 EDITGTTKLVA 1197

Pedant information for DKFZphtes3_1817, frame 2

Report for DKFZphtes3_1817.2

[LENGTH] 1050
 [MW] 117013.72
 [PI] 6.47
 [HOMOL] TREMBL:DMANKY_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,
 complete cds. 2e-45
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13
 [FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
 3e-12
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YDR264c] 3e-12
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08
 [FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]
 3e-08
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
 5e-05
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
 5e-05
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04
 [BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att
 dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12
 [SCOP] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12
 [EC] phosphotransferase 1e-19
 [PIRKW] nucleus 1e-13

[PIKKW] potassium channel 5e-15
 [PIKKW] early protein 2e-13
 [PIKKW] tumor suppressor 1e-09
 [PIKKW] duplication 1e-14
 [PIKKW] tandem repeat 1e-19
 [PIKKW] heterodimer 1e-14
 [PIKKW] potassium transport 5e-15
 [PIKKW] cell cycle control 1e-10
 [PIKKW] serine/threonine-specific protein kinase 1e-19
 [PIKKW] transmembrane protein 5e-15
 [PIKKW] transport protein 5e-15
 [PIKKW] DNA binding 2e-11
 [PIKKW] oncogene 1e-08
 [PIKKW] ATP 1e-19
 [PIKKW] protein kinase inhibitor 1e-09
 [PIKKW] voltage-gated ion channel 5e-15
 [PIKKW] phosphoprotein 4e-38
 [PIKKW] apoptosis 1e-19
 [PIKKW] liver 4e-09
 [PIKKW] integrin binding 3e-16
 [PIKKW] differentiation 2e-12
 [PIKKW] transforming protein 1e-08
 [PIKKW] alternative splicing 1e-40
 [PIKKW] coiled coil 1e-14
 [PIKKW] peripheral membrane protein 2e-38
 [PIKKW] transcription factor 4e-16
 [PIKKW] transcription regulation 2e-16
 [PIKKW] nucleotide binding 5e-15
 [PIKKW] phosphoric monoester hydrolase 1e-12
 [PIKKW] cytoskeleton 8e-39
 [PIKKW] calmodulin binding 1e-19
 [PIKKW] smooth muscle 1e-12
 [SUPFAM] ankyrin 1e-40
 [SUPFAM] death-associated protein kinase 1e-19
 [SUPFAM] ankyrin repeat homology 1e-40
 [SUPFAM] protein kinase homology 1e-19
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07
 [SUPFAM] int-3 transforming protein 1e-08
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38
 [SUPFAM] notch protein 2e-12
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13
 [SUPFAM] rel homology 2e-11
 [SUPFAM] EGF homology 2e-12
 [PROSITE] ATP_GTP_A 1
 [PFAM] Ank repeat
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.05 %

SEQ MALYDELLKNPFYALQKCRPDLCRKVAQIHGIVLPCKGSLSSSIQSTCFESYILIP
 SEG
 lawcB
 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVPILFETTFYNEKEESFSILCIAHPLEKR
 SEG
 lawcB
 SEQ ESSEELAPSDPFSLKTIEDVREFLGRHSEFDRNIASFHRTFRECERKSLRHHIDSANA
 SEG
 lawcB
 SEQ LYTCLQQLLRDHLKMLAKQEAQMNLKQAVEIYVHHEIYNLIFKYVGTMEASEDAAFN
 SEG
 lawcB
 SEQ KITRSLQDLQKQKDIGVKPEFSFNI PRAKRELAQLNKCTSPQOKLVCLRKVVQLITQSPSQ
 SEG
 lawcB
 SEQ RVNLETMCADDLLSVLLYLLVKTEIPNWMANLSYIKNFRFSSSLAKDELGYCLTSFEAAIE
 SEG
 lawcB
 SEQ YIRQGSLSAKPPESEGFCDRLFLKQMSLLSMTSSPTDCLFKHIASGNQKEVERLLSQE
 SEG
 lawcB
 SEQ DHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFPSRDDRGHTPLHVAACVGGASLID
 SEG
 lawcB

36.38 (bits) f: 777 t: 804 Target: dkfzphtes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query *GyTPLHIAARYNNvEMvrILLQHGADIN*
PLH+A++++ ++V+ LL+ +A +N
dkfzphtes3 777 QAVPLHLACQGGHFQVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzphtes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM *GyTPLHIAARYNNvEMvrILLQHGADIN*
G+TPL++A+ ++ E+V LLLQHGGA+IN
Query 810 GNTPLIYACSGGHHELVALLQHGASIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzphtes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query *GyTPLHIAARYNNvEMvrILLQHGADIN*
G+T+LH A+++ +V +V+LLL HGA++
dkfzphtes3 843 GNTALHEAVIEKHVFVVELLLHGASVQ 870

DKFZphtes3_19f19

group: testes derived

DKFZphtes3_19f19 encodes a novel 254 amino acid protein with weak similarity to *S. cerevisiae* protein YFL046w.

The protein contains a RGD cell attachment site.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp

Poly A stretch at pos. 1367, no polyadenylation signal found

```
1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTG CCAGCATTCC TGTGGTGCCA GAACCTACCTT GCCCGAAAAGC
101 CTGTGCGAGA TTTACCCCGT CTTCGCCCTC CTTCCACCGG GAAACTCTTG
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCGCGGG CATTCTCGC CGGCCCTGCG
251 GAGAGAGTTC TCACTACCA CAACCAAGGA GGGATATGAT AGGCCGCCAG
301 TGGATATAAC TCCTTTAGAA CAAGGAAAT TAACCTTTGA TACCATGCA
351 TTGGTTTCAAG ACTTGGAAAC TCATGGATT GACATAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAATCG CTTTATCAAA TGTCAAGCTG GATACTATCT
451 ATAAAGAGAT GGTCACTCAA GCTCAACAGG AAATAACAGT ACAACAGCTA
501 ATGGCTCATT TGGATCTTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAAG
551 TGAATTGCA AATCTGAGAG CAGAGATGA GAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCAACAATA ATGCATGAAA CCAGTCCAGT CAGAGCAGAT
651 AATAAAGTGG ATATCAACTT AGAAAGGAGC AGAGTAACAG ATATGTTTAC
701 AGATCAAGAA AAGCAACTTA TGGAAACAC TACAGATTT ACAAAAAGCG
751 ATACTCAAA CAAAAGTATT ATTTACAGA CCAGTAATAA AATTGACGCT
801 GAAATTGCTT CCTTAAAAAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTTTACTTG CTTGGCAATA GCATTGGGAT
901 TTTATAGATT CTGGAAGTAG TATTAAATGT CATCCTGCTG TGGCTGTTGG
951 CTTCTTAGAA CACCAACCG GGAGAGATT ACTTGAACA TTGTCAGTTG
1001 CAGCAAAAAT TTACTACACA AGATTATTCC AAGTGTATAC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATACT GTGTCTTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGCTTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCTTTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAC CATTCCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATAATTA CATTTTTCTA
1351 TAAATGAAA GATTATTACA ACAAAAAAAA AAAAAAAA AAAAA
```

BLAST Results

Entry HS419346 from database EMBL:

human STS WI-13569.

Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:

human STS SHGC-50338.

Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:

human STS WI-13893.

Score = 1578, P = 1.0e-64, identities = 358/397

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 917 bp; peptide length: 254
Category: similarity to unknown protein
Classification: no clue
Prosite motifs: RGD (15-18)

1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTTT KEGYDRRPVD
51 ITPLEQRKLT FDTHALVQDL ETHGFDKTQA ETIVSALTAL SNVSLDTIYK
101 EMVTQAQQEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMKIELDQ
151 VKQQLMHETS RIRADNKLDI NLEERSVTDM FTDQEKQLME TTTEFTKKDT
201 QTKSISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLALALGFY
251 RFWK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19f19, frame 3

SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME
I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 138, P = 5.4e-08

>SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.
Length = 211

HSPs:

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09
Identities = 34/121 (28%), Positives = 67/121 (55%)

Query: 70 LETHGFDKTAETIVSALTALSNVSLDTIYKEMVTQAQQE-ITVQQLMAHLDAIRKDMVI 128
LE G+ AETI + + + + +L + K + +A+QE + + QQ L IRK +
Sbjct: 46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAKQESVSFQQKRTFLQ-IRKYLET 104
Query: 129 LEKSEFANLRAENNEKMKIELDQVKQQLMHETSRIKADNKLDINLERSRVTDMDQEKQL 188
+E++EF +R ++K+ E+++ K L + ++ +L++NLE+ R+ D T + +
Sbjct: 105 IEENEFDKVRKSSDKLINEIEKTKSSLRDVKLTALSEVRLNLNLEKGRMKDAATSRNTNI 164
Query: 189 ME 190
E
Sbjct: 165 HE 166

Pedant information for DKFZphtes3_19f19, frame 3

Report for DKFZphtes3_19f19.3

{LENGTH} 254
{MW} 29505.73
{pI} 6.99
{HOMOL} PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YFL046w] 8e-12
{PROSITE} RGD 1
{KW} TRANSMEMBRANE 1
{KW} LOW_COMPLEXITY 5.12 %
{KW} COILED_COIL 11.02 %

SEQ MNSRQAWRLF LSQGRGDRWVSRPRGHFSPALRREFFTTTTKEGYDRRPVDITPLEQRKLT
SEG
PRD ccchhhhhhhhhccccccccccccchhhhhhhheeeccccccccccchhhhhhhcc
COILS
MEM
SEQ FDTHALVQDLETHGFDKTAETIVSALTALSNVSLDTIYKEMVTQAQQEITVQQLMAHLD
SEG
PRD chhhhhhhhhhhccccchhh
COILS

PCT/IB00/01496

Prosites for DKFZphtes3_19f19.3

(No Pfam data available for DKFZphtes3_19f19.3)

DKFZphtes3_19j17

group: testes derived

DKFZphtes3_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp

Poly A stretch at pos. 2740, no polyadenylation signal found

```
1 ATTCTCAGCC AAATTTTTTT ATTTTGTGCA GAATCAGTGT GCAAGGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTTTTTGT GTTTAGTGTG ATTTGTTATC
101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTTTGTG TGAGAAACTT
151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAA TTTTAATTTT GTAAGTTCAT
201 AAAAATTTAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
251 GGTTCATATT CTTTCTCTTT TTTTTTGGGG GGTGTATGTT TTACAGGAC
301 TTAAGTATTC ATCGAAGAGT CACCCACGTA GCGGTGATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAATAAATA TGTTCGGAG
401 ATCTGATAGT CCTGAAACAA AATACAGTGA CAGCACAGCT CACAGTAAG
451 CCAAAAATGT GCATACCTAC AGAGTTAGAG AGAGCGATGG TGGGACCACT
501 TACTCTCCAC AAGAAAATTC ACACACCCAC AGTCTCTTTC ATAGTTCAAA
551 TTCACATTC TCTAATCCAA GCAATAACCC AAGCAAAACT TCAGATGCAC
601 CTTATGATTC TCGAGATGAC TGGTCTGAGC ATATTAGCTC TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG ACAGAGAAGT TCACATGGG AAAAACCAAA
701 AGACTGGCTT GAAGAGAAC AGAGACAAA AGAAGCAAA AGATGGCAG
751 TCACAGCTT CCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGGAAAT GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTTT GTCTCAAACA AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CAGTACAGC ACCCCATCAA ACCAGTGGT CATCCAACGT CTACCCCAAG
1001 CACTGTTCTT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTGTA TGCTAATGGA GCATCTACTT TATCAAACCT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAACAGAAA AGAAAAGAAT CTACATCAGG
1151 AGACAAACCC GTATCACATT CTTGCACAC TCCTTCCAGC TCTTCTCCCT
1201 CTGGACTGAA CCCACATCT GCACCTCCAA CATCTGCTTC AGCGTCCCT
1251 GTTCTCCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTCTTTC CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTC TCAAGTGTCT CAGCTCTCTA
1501 CACAAGATAT CCTCTTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAAT GGAAGTGAA AGGGTCACCT TGTGAGAAAG CTGATAACAA
1601 GCAGGAATGC CTTGCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCCTCTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAATAAAGG ATGCCAGGCC ATCTAATCAG
1751 TCTCCAGTGT CTTTAACATC TGATGCGTCA TCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACCTAACAC AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTTAGTAC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTACCA GTCAGCCACA CAGCAGCCTG TAACGTGCTGA
1951 CAAGCAGCAA GGTATGAAC CTGTCTCTCC TCGAAGCTTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAATG CAACAGTGTG ACCACAGAA TCTTCTGCCC GATCCACGTG
```



```
2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGCGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTGTGAC
2251 TGAATTAAAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATCAAGCAA
2301 CTTTGCGAGA GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCT
2351 GAAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCC AATCTTAACA
2451 TTTTGTAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGATC AATTCAGGG AAAGATACAA
2551 GATTGATTG TAAAACCCCTT GAAATGTAGA TTTCTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTTG TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA GTTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA
```

BLAST Results

Entry AC005876 from database EMBLNEW:
Homo sapiens chromosome 10 clone CIT987SK-118815 map 10p11.2-10p12.1,
complete sequence.
Score = 2130, P = 0.0e+00, identities = 426/426
12 exons matching Bp 492-2740

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209
Category: questionable ORF
Classification: no clue

```
1 MSLTSDASSP RSYVSPRIST PQNTVPKPK LISTPPVSSQ PKVSTPVVKQ
51 GPVSQATQQ PVTADKQGGH EPVSPRLQR SSQSPSPGP NHTSNSSNAS
101 NATVVPQNSS ARSTCSLTPA LAHFSENLI KHVQGWPDH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLRVCEIQ ATLREQRILE LRQIQELEK
201 LKNQNSFMV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436
Category: similarity to unknown protein
Classification: unclassified
Prosites motifs: WW DOMAIN_1 (90-116)
WW_DOMAIN_1 (90-116)

```
1 MRDAGDPSPP NKMLRRSDSP ENKYSDSTGH SKAKNVHTR VREDDGTSY
51 SPQENSHNHS ALHSSNSHSS NPSNNPSKTS DAPYDSADDW SEHISSSGKK
101 YYNCRTEVS QWEKPKEWLE REQROKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSOTS RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPOSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAATV
351 QASLQSIHK FLTAGPSAFN ITSLSISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSLC QKADKQCECL VWNGSIMVQR LLQPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 3

TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A, N = 1, Score = 144, P = 1.8e-09

>TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A
Length = 120

HSPs:

Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09
Identities = 30/67 (44%), Positives = 43/67 (64%)

Query: 90 WSEHISSSGKKYYNCRTEVSQWEKPKEW-LEREQRQKEANKMAVNSFPK---DRDYRRE 145
W+E +SSSGK YYN +TE+SQW+KP EW E +++ K VN P+ DR Y
Sbjct: 11 WTEQMSSSGKMYYNKKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query: 146 VMQATATS 153
+ Q +++S
Sbjct: 70 IGQLSSSS 77

Pedant information for DKFZphtes3_19j17, frame 2

Report for DKFZphtes3_19j17.2

[LENGTH] 209
[MW] 22873.85
[pI] 9.95
[KW] All_Alpha
[KWL] LOW_COMPLEXITY 13.40 %

SEQ MSLTSDASSPRSYVSPRISTPQTNTVPIKPLISTPPVSSQPKVSTPVVKQGPVSQSATQQ
SEG
PRD ccc
SEQ PVTADKQGGHEPVSPRSLQRSSQSPSPGPNHTSNSSNATVVPQNSSARSTCSLTPA
SEG
PRD ccc
SEQ LAAHFSENLIKHHVQGWADHAERQASRLREEAHNMGTIMHSEICTELKNLRLVVRVCEIQ
SEG
PRD hhhhhhhcc
SEQ ATLREQRILFLRQQIKLEKLKNQNSFMV
SEG
PRD hhh

(No Prosite data available for DKFZphtes3_19j17.2)

(No Pfam data available for DKFZphtes3_19j17.2)

Pedant information for DKFZphtes3_19j17, frame 3

Report for DKFZphtes3_19j17.3

[LENGTH] 436
[MW] 47716.62
[pI] 8.71
[HOMOL] TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A 6e-08
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPR152c] 6e-04
[BLOCKS] BL01159 WW/rsp5/WWP domain proteins
[PROSITE] WW DOMAIN 1 2
[PFAM] WW/rsp5/WWP domain containing proteins
[KW] All_Alpha
[KWL] LOW_COMPLEXITY 22.48 %

```

SEQ      MRDAGDPSPPNKMLRRSDSPENKYS DSTGHSKAKNVHTRVREDDGTSYSPQENSHNHS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      ALHSSNSHSSNPSNNPSKTS DAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKEWLE
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh

```

```

SEQ      REQRQKEANKMAVNSFPKDRDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQT
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      RHNDRDYRLPRAETHSSSTPVQHPKIPVHPTATPSTVPSSPFTLQSDHQP KKSFDANGA
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      STL SKLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPSTSSASGLNPTSAPPTSASAVPV
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SPVPQSPIPFLQDPNLLRQLPALQATLQLNNSNVDISKINEVLTAAVTQASLQSI IHK
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      FLTAGPSAFNITSLISQAALSTQDIP LHEGIQMERDTHRSKWEVKSLCQKADKQEQECL
SEG      .....
PRD      hhccccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      VWNGSIMVQRLQPSG
SEG      .....
PRD      eecchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Prosite for DKFZphtes3_19j17.3

PS01159	90->116	WW_DOMAIN_1	PDOC50020
PS01159	90->116	WW_DOMAIN_1	PDOC50020

Pfam for DKFZphtes3_19j17.3

```

HMM_NAME      WW/rsp5/WWP domain containing proteins
HMM            *LPsGWEeHWDPsGRpWYYWNHETkTTQWEp*
               + **W EH** SG* YY+N T* +QWE+P
Query         86 SADDWSEHISSSGKK-YYNCRTEVSQWEKP 115

```

DKFZphtes3_lcl

group: signal transduction

DKFZphtes3_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```
1 GCGAAGTGAA GGGTGGCCCA GGTGGGCGCA GGCTGACTGA ATGTATCTCC
51 TAGCTATGGA CTAATAATA CATGGGGGGA AATAACAAG TATTCATGAG
101 GGTGAAATG TGACCCAGCA GGAAATATC AACTATTTTC AATTGACGTT
151 GAATAGGATG AGTCATGAA TTTAAGTGAT TTAAGTGAAG TTATACTACT
201 GGTAGATAGA AGAGCTAAG AAAGATGGAT ACTATGATCG TGAATGTGCG
251 GAATCTGTTT GAGCAGCTTG TGCGCCGGGT GGAGATTCTC AGTGAAGGAA
301 ATGAAGTCCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCGGTAAA
351 AAGTGGCAGA GGAAGTACCA TGAGCTGGGG AAATACAAGG ATCTTTTGAT
401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTTAAGCTG AAGCATGCAC
451 GTAATCAGGT GGATGTAGAG ATCAACCGGA GACAGAGAGC TGAGGCTGAC
501 TGCGAAAGGC TGAACGACA GATTCAGCTG ATTCGAGAGA TGCTCATGTG
551 TGACACATCT GGCAGCATTG AACTAAGCGA GGAGCAAAAA TCAGCTCTGG
601 CTTTCTCAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAAAAG
651 CTATCAACCA TTGATGAATC TGGTTCCATT TTATCAGATA TCAGCTTTGA
701 CAAGACTGAT GAATCACTGG ATTTGGGACTC TTCTTTGGTG AAGACTTTGA
751 AACTGAAAGAA GAGAGAAAAG AGGGGCTCTA CTAGCCGACA GTTTGTTGAT
801 GGTCCCGCTG GACCTGTAAA GAAACTCCTT TCCATGGGCT CTGCACTAGA
851 CCAGGGGAAT CAATCCATAG TTGCAAAAC TACAGTGACT GTTCCCAATG
901 ATGGCGGGCC CATCGAAGCT GTGTCCACTA TTGAGACTGT GCCATATTGG
951 ACCAGAGGCC GAAGGAAGAC AGGTACTTTA CAACCTTGGA ACAGTGACTC
1001 CACCCTGAAC AGCAGCGACC TGGAGCCCAAG AACTGAGACA GACAGTGTGG
1051 GCAGCGCACA GAGTAATGGA GGGATGCGCC TCGATGACTT TGTTCCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTTCCA TGTGGAAGC GGATAAAATT
1151 TGGCAATTA TCTCTAAGT GTGAGACTGT TCGTGTGGTC TCTCATCCAG
1201 AATGTGCGGA CGCTGTCCC CTTCCCTGCA TTCTACCCCT GATAGGAACA
1251 CCTGTCAAGA TTGGAGAGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCATTG TGTAAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCGCACA
1401 GTAAAAGAGC TGAAAGAGAA ATTCCTCAGA GTGAAAACCTG TACCCTCTCT
1451 CAGCAAGTGT GATGATATCC ATGCTATCTG TAGCCCTCTA AAAGACTTTT
1501 TTCGAAACCT CAAAGAACCT CTTCTGACCT TTGCGCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAACCTG CCCAGGCCAA CAGGGACACA TTAGCTTTCC
1651 TCATGATTCA CTGACAGAGA GTGGCTCAGA GTCCACATAC TAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGGCCCT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCCTGCTT TCCTTGCCCT TGGAGTATTG GAGTCAGTTC
1851 ATGATGTGGG AGCAAGAGAA CATTGACCCC TCATCATGTA TTGAAAACCT
1901 AAATGCCTTT TCAACACCAC AGACACCAGA TATTAAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCCTGAA CATCAGCTTC TCAAGACTCC TTCTCTAGT
2001 TCCCTGTGAC AGAGAGTCCG TTCCACCCTC ACCAAGAAC CTCTAGATT
2051 TGGGAGCAAA AGCAAGCTCT CCACATAACCT AGGACGACAA GGCAACTTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCTGTT ACTTCCAGC
2151 ATTGACTGAC TATAAGAAAG GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTAATCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTCATGAG GGTTTTATTA AAATATATA TATCTCCCTC TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TCTTTTTTGG GGAAGGGGGT TATTGTTCTT TTGGCTTAAA GCCAAATGCT
2451 GCTCATAGAA TGATCTTTCT CTAGTTTCAT TTAGAAGCTA TTCCGCTGAG
2501 ACAATGACAG AAACCTTACC TATCTGATAA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTAGGAT
```

```

2601 GCCTCCTTCT AAGAACCCAGA AGTTCTCATT CCCCATTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTTGGATGA TTAGGTCTTT AATAGTGTGG
2751 AGTGGCACAA CTTGTAAAT GTGAAAGTAC AACTCGTATT TATCTCTGAT
2801 GTGCCGCTGG CTGAACCTTG GGTTCATTTG GGGTCAAGCC CAGTTTCTCT
2851 TTTAAATTTG AATTCACTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCACTGGAGC CCAACTTCTA AAGGTCAATA TATCATCAAT TGGCATCCCA
2951 ACTAACCAATA AAGAGTAGGC TATAAGSGAA GATTGTCAAT ATTTTGTGGT
3001 AAGAAAAGCT ACAGTCATTT TTTCTTTGCA CTTTGGATGC TGAATTTTCT
3051 CCCATGGGAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTGA
3101 AAATTATTCT TAATGCTCTG TAAACGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCTATTG ACCCTTATCT GTAAACACC TATTGGGAT AATATTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAA

```

BLAST Results

Entry U82984 from database EMBLEST:
Homo sapiens DRES 56 mRNA sequence.
Score = 8775, P = 0.0e+00, identities = 1757/1758
matches 3' end

Medline entries

93074974:
Developmental regulation and neuronal expression of the mRNA of rat
n-chimaerin, a
p21rac GAP:cDNA sequence.

93024458:
A Drosophila rotund transcript expressed during spermatogenesis and
imaginal disc
morphogenesis encodes a protein which is similar to human Rac
GTPase-activating
(racGAP) proteins.

Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632
Category: similarity to known protein

```

1 MDTHMLNVRN LFEQLVRRVE ILSEGNEVQF IOLAKDFEDF RKKWQRTDHE
51 LGKYKDLLMK AETERSALDV KLKHARNOVD VEIKRRORAE ADCEKLERQI
101 QLIREMLMCD TSGSIOLSEE QKSALAFILNR GOPSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLOWDSS LVKTFKLKRR EKRRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNEISIVAKTT VTVPNDDGPI EAVSTIETVP YWTRSRRTKG
251 TLOPWNSDST LNSRQLEPRT ETDSVGTQPS NGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKCRDCR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFSQTSFP MIPSIVVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LKDFLRLNK EPLLTFLNR AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPTMLQD IKROPKVVER LLSLPLEYWS QFMVVEQENI
551 DPLHVIENSN AFSTPQTPDI KVSLLGPVTT PEHQLLKTFS SSSLSQVRVS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK

```

BLASTP hits

Entry CEK08E3_4 from database TREMBLNEW:
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit
fly (Drosophila melanogaster) (fragment)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit
fly (Drosophila melanogaster)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539.1 from database TREMBL:
gene: "rotund"; product: "InracGAP"; *Drosophila melanogaster* rnracGAP
(rotund) gene, complete cds.
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:
N-chimerin - rat
Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKFZphtes3_1c1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_1c1, frame 3

Report for DKFZphtes3_1c1.3

[LENGTH] 632
[MW] 71026.84
[PI] 9.08
[HOMOL] PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -
fruit fly (*Drosophila melanogaster*) 2e-46
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]
2e-11
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YOR127w] 5e-09
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08
[FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08
[BLOCKS] BL00479B Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS] BL00479A Phorbol esters / diacylglycerol binding domain proteins
[SCOP] dlphwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo)] 1e-55
[SCOP] dlrgp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 1e-49
[PIRKW] breakpoint cluster region 1e-19
[PIRKW] transmembrane protein 7e-08
[PIRKW] brain 3e-22
[PIRKW] alternative splicing 1e-19
[PIRKW] P-loop 2e-25
[SUPFAM] CDC24 homology 3e-22
[SUPFAM] bcr protein 3e-22
[SUPFAM] myosin motor domain homology 2e-25
[SUPFAM] pleckstrin repeat homology 4e-10
[SUPFAM] LIM metal-binding repeat homology 2e-09
[SUPFAM] protein kinase C zinc-binding repeat homology 5e-29
[PROSITE] MYRISTYL 6
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 9
[PROSITE] ASN_GLYCOSYLATION 1
[PROSITE] DAG_PE_BINDING_DOMAIN 1
[PFAM] Phorbol esters / diacylglycerol binding domain
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 2.22 %
[KW] COILED_COIL 8.54 %

SEQ MDTMMLNVRNLFELVRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK
SEG
COILSCCCCCCCCCCCC
1rgp-.....
SEQ AETERSALDKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSIQLSEE
SEG
COILS CC
1rgp-.....
SEQ QKSALAFNLRGQPSSSNAGNKLSTIDESGSILSDISFDKTDSELDWSSLVKTFKLKKR
SEG
COILS

```

1rgp- .....
SEQ      EKRRSTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDDGGPIEAVSTIETVP
SEG      .....
COILS    .....
1rgp- .....

SEQ      YWTRSRRTGTGLQPNWSDSTLNSRQLEPRTEDSVGT PQSNGGMRLHDFVSKTVIKPESC
SEG      .....
COILS    .....
1rgp- .....

SEQ      VPCGKRIRKFGKLSLKRDCRVVSHPECRDRCLPCIP TLIGTPVKIGEGMLADFSQTSP
SEG      .....
COILS    .....
1rgp- .....

SEQ      MIPSIVVHCVNEIEQRLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG      .....
COILS    .....
1rgp-    .CCHHHHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHCCCCG-GGCCCCHHHHH

SEQ      LLKDFLRNLKEPLLTFRLNRAFMAAEITDEDNSIAAMYQAVGELPQANRDTLAFIMHL
SEG      .....
COILS    .....
1rgp-    HHHHHHHHTTTTTTGGGHHHHHTTTT-CGGGHHHHHHHHHHHCCCHHHHHHHHHHHH

SEQ      QRVAQSPTKMDVANLAKVFGPTIVAHAVPNPDVMTLQDIKRQPKVVERLLSLFLEYWS
SEG      .....
COILS    .....
1rgp-    HHHHHHHHCCCHHHHHHHGGGCC.....

SEQ      QFMVQEENIDPLHVIENSNAFSTPQTPIKVSLLGPVTTPEHQLLKTSPSSSSLSQRVRS
SEG      .....
COILS    .....
1rgp-    .....

SEQ      TLTKNTPRFGSKSKSATNLGRQGNFASPLK
SEG      xxx.....
COILS    .....
1rgp-    .....

```

Prosites for DKFzptes3_lcl.3

PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00004	141->145	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	182->186	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	246->250	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	174->177	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	313->316	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	595->598	PKC_PHOSPHO_SITE	PDOC00005
PS00005	606->609	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	270->274	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	387->391	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	410->414	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	489->493	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00007	46->55	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->385	TYR_PHOSPHO_SITE	PDOC00007
PS00008	131->137	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	377->383	MYRISTYL	PDOC00008
PS00008	388->394	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00009	303->307	AMIDATION	PDOC00009

WO 01/12659

PCT/IB00/01496

PS00479 287->336 DAG_PE_BINDING_DOMAIN PDOC00379

Pfam for DKF2phtes3_lcl.3

HMM_NAME	Phorbol esters / diacylglycerol binding domain		
HMM	*HrFmrHTFrqPTWCDHCgeFIWGWgKQGYOCQnCgMNCHKRCHelVPmm		
	H+F+ +T + P +C CG +I +GK ++C +C+++ H +C+ + P		
Query	287	HDFVSKTVIKPESCVPCGKR1-KFGKLSLKCRDCRVVSHPECRDRCLP	334
HMM	C*		
	C		
Query	335	C	335

DKFZphtes3_lgl3

group: intracellular transport and trafficking

DKFZp DKFZphtes3_lgl3 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cpl51 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cpl51"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cpl51 shows haploid-specific transcription!

testis or haploid-specific transcription

Sequenced by DKFZ

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```
1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
51 CATTCTCTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
101 AGCTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGGCGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTCTGAG GAGTCAGAGG TGGAATTTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAGAAAAAAT
401 TGCTGGTCTT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGACAGCT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCTAGACGA AGCAGACTTC
501 CGACCTGCTT CTCTCGACAC ATCACTGCAC ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAATCACA ACGAGAACAC AGGGGAGAAAG
601 CTCCATTGGG CGCAGGAGCA ACTCGCTTGG GCCGGGGACA AGATCGCCTC
651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAG TCTTCCCTGA
701 GCAACATCGA GTTACTAGAA TGCCAAAGTA AGATGTTGCA GGGGGAACCTC
751 GGCGGGATCA TGGGTCAGGA GCCTGAGAAC AAGGGTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCTCT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCAACAGGA TGATCTCATT
901 CAAGAACTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
951 TGAAAAGGCT TTGATAAAAC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCTCTCTAGC TCCTCAGGAG AGTGTAAGA CATCAAAAAG
1051 ATACTGAAGC ACTTGCAAGG GCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACCTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TGGGGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCCTGCAGT GCCGGCTGCA GGAGCTGCAG CTGGAGTTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCTC CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCATGTC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCTGCGAGGC TGAGGTCAGC AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCAGTGCCA GGCTGTCAAC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAAGG GTCTCTCTCT GGACAAGCAG AAGGACAGCA CCATCCAGGA
1701 ACTACAGAGA GAACTTCAGA TGCTGCAGAA GGAGTCTCTG ATGGCTGAGA
1751 AGGAACAAAC CTCACACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAATAATCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCACAGG GAGCAAGGCT CCATCAAAAT CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTGGAAGCTT
2051 TGCGGCAGGA ATTTAAAAAG AAAGACAAGA CTTTGAAGA GAATTCAGGA
2101 AAGTTGGAGG AAGAAAAATG GAATCTCCGA GCAGAGCTAC AGTGTGTTT
2151 TACACAACCTG GAATCTCTCT TCAACAATAA CAACCCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGACAGGCC AGCTGGAGAA ACCTCTGCAG AAGGAGAGC ACTATCTCCA
2301 GACTACATC ACCAAAGAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CCTGCCAGGA TGACCTGACA CAAGCCCTCG AGAAGTCAA TCAGGTGACC
2401 TCAGAGACAA AGAGCTGCA GCAAAGCTTG ACACAGACC AAGAGAAGAA
```

```

2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGCACG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGAGGTTT CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTGC CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCGAAAAGGA
2901 CCACCTCCAC AGTGTAAATG TCCACTTGCA GCAGGAAAC AAGAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAGGCC TAGGCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGCGGCACC TTGGGCTGGA AGGGGTGCCC CAGGATATG GGTCAAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCTCATAC
3151 TGCTAGAATC CACATCTAGC CCTGAGCAGC ATTCCACGG GTGTTTCTTC
3201 AGAGGACAGT GAGTTCCAG CCTCCCTCT CTCTTGACCT GGATCAGCTG
3251 TTACAGGAGT ATATCAGCGT CCCAGCCTAT TTGCAAGAC ACTAACTTTT
3301 GTTGAGTTTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCAATCTCCAG GCCCAAACTC TGAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA

```

BLAST Results

Entry AC004682 from database EMBLNEW:
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete
sequence.
Score = 1291, P = 0.0e+00, identities = 265/272

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007
Category: similarity to known protein
Prosite motifs: LEUCINE_ZIPPER (83-105)
LEUCINE_ZIPPER (90-112)
LEUCINE_ZIPPER (97-119)
LEUCINE_ZIPPER (104-126)
LEUCINE_ZIPPER (403-425)
LEUCINE_ZIPPER (410-432)
LEUCINE_ZIPPER (918-940)

```

1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQQLKKKL LVLQOELEFH
101 TEELQTSYYS LRQYSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
151 ENTGEKLHLA QEQLALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
201 MLQELGGIM GOEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVMQKVS
251 QODDLIQELR NKLACSNALV LEREKALIKL QADFASCTAT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQRRNIMKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLLKKD
401 KFLQEKDEML QELEKKLTQV QNSLLKKEKE LEKQQCHATE LENTVKEAQ
451 DKSKEAECKA LQAEVQKLKN SLEBAKQQR LAAQQNAQCK EEAALAGCHL
501 EDTQRKLQKG LLLDKQADT IQELQRELQW LQKSSMAEK FQTSNRKRVE
551 ELSLESEAL RKLNSDKKX RQLQKTVAEQ DMKMDNMLDR IKHQHREQGS
601 IKCKLEEDLO EATKLEEDKR EQLKKSKEHE KLMGELEAL RQEFKKDKDT
651 LKENSRLKEE ENENLRAELQ CTSQLESSL NKYNTSQOVI QDLNKEIALQ
701 KESLSLQAO LKALQKEXH YLQTTITKEA YDALSRSAA CQDDLTOALE
751 KLNHVSETK SLQSLTOTO EKKAQLEEEI IAYEERMKKL NTELRLRGF
801 HQSELEVHA FDKLEEMSC QVLQWQKQHQ NDLKMLAAKE EQLREFQDEM
851 AALKENLLED DKEPCCLPQW SVPKDTCRLY RGNDQIMTNL EQWAKQKVA
901 NEKLGQLRE QVNYIAKLSG EKHDLHSMVY HLQENKKLK KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGTLGWK GLPQDMGQRM DLTXYIGMPH
1001 CPGSSYC

```

BLASTP hits

Entry H5417401_1 from database TREMBL:
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete

cds.
Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862

Entry SCINTANA_1 from database TREMBL:
Saccharomyces cerevisiae integrin analogue gene, complete cds.
Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897

Entry HS6802_2 from database TREMBL:
gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC
6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain,
ESTs, CA repeat, STS and GSS.
Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028

Entry AF092090_1 from database TREMBL:
product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

Alert BLASTP hits for DKF2phtes3_lg13, frame 1

TREMBL: HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin,
N = 1, Score = 411, P = 4.4e-34

TREMBL: HS417401_1 product: "trans-Golgi p230"; Human trans-Golgi p230
mRNA, complete Cds., N = 1, Score = 411, P = 4.5e-34

TREMBL: SCINTANA_1 Saccharomyces cerevisiae integrin analogue gene,
complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL: HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin
Length = 2,185

HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34
Identities = 212/816 (25%), Positives = 420/816 (51%)

Query: 145 EMGNHNEN-TGEKHLAQEQALAGDKIASLERSLNLYRDKYQSSLSNIELECCQVKMLQ 203
+M + E+ G L +EQL ++ +ERSL+ YR KY ++ ++L+ + K LQ
Sbjct: 219 DMDSEADLVGNSDSLNEQLI---QRLRRMERSLSSYRGKYSLVYAYQMLQREKKKLQ 175

Query: 204 GELGGINGQEPENKGDHSHKVIYTSPCMIQEOHETQKRLSEVMQ-KVSQQDDLIQELRNK 262
G I+ Q D S RI +Q Q+ +K L E + ++D I L+ +
Sbjct: 176 G----ILSQSQ----DKSLRRIAELREELQMDQAKKHLQEEFDASLEEKDQYISVLQSQ 227

Query: 263 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKHLQE 313
++ + + ++ K L +L+ A P S E ED K L+ LQ+
Sbjct: 228 VSLKQRLRNGPMNVDLKPLPQLEPQ-AEVFTKEENPESDGEPPVEDGTSVKTLETQQ 286

Query: 314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQRRNIMKDMMLKLELDLHGLREETS 366
+ Q C ++ ++ L E EA+ EQ ++++ K+ DLH + E+T
Sbjct: 287 RVKRQENLLKRCETIQSHKEQCTLLTSEKALQEQDLERLQELEKIK-DLH-MAEKTKL 344

Query: 367 HIERKDKDITILQRLQELQLEFFTETQKLTLLKKDKFLQEKDEMLQELEKKLTQV--QNSL 424
+ +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q
Sbjct: 345 ITQLRDAKNLIEQLE-QDKGMVIAETKR---QMHTLEMKEEIAQLRSRIKQMTTGEE 400

Query: 425 LKKEKELEKQQCMATELEMTVKEAKQDKSKEADCKALQAEVQKLKNSLEEAKQERLAAQ 484
L+++KE + + ELE + K+ K+EA K L+AE+ + ++E+ ++ER++ Q
Sbjct: 401 LREQKE-KSERAAFEELKALSTAQ--KTEARRK-LKAEMDEQIKTIKTSEEERISLQ 456

Query: 485 QA-AQCKEEAA-LAGCHLEDTQKQLQGLLLDKQKADTIQELQRELQMLQKSSMAEKEQ 542
Q ++ K+E + E+ KLQK L +K+ A QEL ++LQ ++E E+ +
Sbjct: 457 QELSRVQEVVDVMKKSSEEQIAKLQK--LHEKELARKEQLTKKLQTRERE--FOEQMK 512

Query: 543 TSNRKRVEELSLELSEALRKLENSDKERQLOKT--VAEQDMKNDMLDRIKHQHREQGS 600
+ K E L++S+ + E+ E+ +LQK + E + K+ D+ +
Sbjct: 513 VALEKSQSEY-LKISQKEQESLAELELQKKAILTESENKLRLDQEAETVTRILE 571

Query: 601 IKCKLEEDLQEA TKLLED----KREQLKKSKEHEKLMEG--EELALR-QEFKKKDKTL 651
++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L
Sbjct: 572 LESSLEKSLQENKNQSKDLAVHLEAKNKHKNKEITVMVEKHKTELESLEKHKQDALWTEKL 631

Query: 652 KENSRLKEEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706
+ ++ + E E LR + C + E+ L +K Q I+++N++ + +++ L S
Sbjct: 632 QVLKQQYQTEMKLRK---CEQEKETLLKDKKEIFQAHIEEMNKTLEKLDVKQTELES 688

Query: 707 LQAQLDKALQKEKHLYQT--TITKEAYDALSRKSAACQDDLTQALEKLNHVTSKSLQQ 764
L ++L + L K +H L+ ++ K+ D + ++ A D+ Q V S K +

Sbjct: 689 LSSELSVL-KARHKEEELSVLKDQTKMKQLEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGFHESELEVHAFDKKLEEMSCQVLQ 824
S+ +T+ KA L+++I E +K+ + L++ + + E ++ + +L++ S ++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLKKERDKHLKEHQAHVENLEADIKRSEGELOQASAKLDV 802

Query: 825 WQKQHNDLKMALAAKEEQLEFQEEAALKENLLEDDKEPCCLPQW----SVPKDT-C-R 878
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + + K D C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLOQKLLDLETERILTRKQVAEVAQKKDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIAKLS-GEKDLHLSVMVHLQEQNK 937
L Q+ ++Q RQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKIQVDLMQQLKQNSEMEQKVSILT--QV-YESKLEDGNKEQEQTKQILVEKENM 912

Query: 938 KKK-KEIEEKKMAENTRLCTK 958
L+ +E ++K+++ +L K

Sbjct: 913 ILQMREGQKKEIEILTQKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLNS-KLL-SLQLDIKNLHDVCKRQKTLQDN-QLCM-----EEM 51
K+E E D E V S K L +LQ +K ++ KR ++T+Q + + C +EA+

Sbjct: 260 KEENPESDGEPPVEDGTSVKTLETQQRVQRQENLLKRCCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSDKKQAQALAFESEVEFGSSKQCHLRQ---LQOLK--KKLLVLQLEFHTTEELQ 105
D++ + ++ + + LR ++QL+ K +++ + + H E L+

Sbjct: 320 QEQLDERLQLEKIKDLHMAEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRQMH-ETLE 378

Query: 106 TSYSYLRQYSILEKQTSDDLVLHHCKLKEDEVILYEEEMGNHNTGEKHLAQEQL- 164
+ Q +S +++ T+ L K K + E E +T+K A+ +L

Sbjct: 379 MKEEIAQLRSRIKQMTTQGEELREQ-KEKSERAAFEELKAL--STAQTEEARRKLK 434

Query: 165 ALAGDKIASLERSLNLRYDKYQSSLSNI--ELLEQVQMLQELGGIMGOEPENKGDHKS 222
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 AEMDEQIKTIKTSEERISLQLESRVKQEVVDVNMKSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYSPCMIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNALVLEREKALIKLOA 282
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 EQELTKKLQRE-REFEQMKVALEK-SQSEYL--KISQKEQESLAE--LELQK 544

Query: 283 DFASCTATHRYPPSSSECEEDIKILKHLQEQKDSQCLHVEEYQNLVLDLVELEAV-SE 341
A T + +E E + + L+ + +E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKRLDQLEAETRYRILELESSLEKS---LQENKQSKDLAVHLEAKNK 600

Query: 342 QRRNIMKMKLELDLHGLREETSAMIERKDKDITI-LQCRLEQLQLEFTETOKLTLKKD 400
+ I + K + +L L+ + A K + + Q +++L+ E +K TL KD

Sbjct: 601 HNKETVMVEKHTELESKHHQDALWTEKLVQKQYQTEMEKLR-EKCEQEKETLLKD 659

Query: 401 K-----FLQEKDEM-LQLEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKS 453
K ++E +E L++L+ K T+++ SL + E+ K + E +E+V + + DK

Sbjct: 660 KEIFQAHIEEMNEKTEKLDVKQTELE-SLSELSVLKARHKEE-ELSVLKDQTKRM 717

Query: 454 K-EAECKALQAEVQKLKNSLEEAKQERLAAQQAQC-KEEAALAGCHLEDQTKLQKGL 511
K E E K + + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKKERDKHLKEHQ 776

Query: 512 L-LDKQKADTIQELQRELQMLQKSSMAKEQTSNRKRVEELSLELSEALRKLNSDKEK 570
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGELOQASAKLDVFQSYQSATHEQTKAYEEQLAQLOQKLLDLETER 835

Query: 571 RQLQKTVAEQDMKMDM---LD--RIKHQHQEGSIK--CKLEEDLQATKLLDKREQL 623
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVAQKKDVCTELDAHKIQVDLMQQLKQNSEMEQKVSILTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMEGELEALRQEFKKDKTLKENSRLKEENENLRAELQCCSTOLESSN 681
K +E K L+E L+ +K K ++ +KL + +++ + T+ ++

Sbjct: 896 NKEQEOTKQILVEKENMILQMREGQK-KEIEILTQKLSAKEDSIHILNEEYETKFNQEK 954

Query: 682 KYNTSQOVIQDLNKEIALQKESLMSLQALDKALQEKHYLOTITKEAYDALSRKSAAC 741
K Q +++ + + K+ L+ +A+L K L E L+ + ++ ++A + A

Sbjct: 955 KMEKVKQAKEMOETL---KKLLDQEAKLKEL--ENTALELSQREKQFNKMLEMAQA 1009

Query: 742 QD-DLTQALEKLNHVSTETKSLQOSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGF 800
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQKEQIE-SLTVHRR--ELNDVISIWE---KKLNQAEELQEI 1061

Query: 801 HQSELEVHAFDKKLEEMSCQVLQW--QKQHNDLKMALAAKEEQLEFQEEAALKENLL 858
H E+++ ++++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLQEKQEVAEKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTLNQLQELK 1116

Query: 859 EDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQ--WAKQKQVANEKLGQNLREQVNYI- 915
+ L Q K L + + +L++ + ++Q V + L + + +V+ +
Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLQEQVLKMLAEEDKRRKVELT 1175

Query: 916 AKLSGEKDLHLSVMVHLQQENKKL-KEIEEKKMAE 951
+KL + S+ ++ NK L+ K +E KK+ E
Sbjct: 1176 SKLKTDEEFQSLKSSHEKSNKSLKEDKSLFKKLEEE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26
Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLNSKLLSLQDIKNLHDVCKRQKTLQDNQLCMEAMNSSHDKKQAALAFESE 69
+E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +
Sbjct: 560 QEAETRYTRILESSLEKSLQENKQKDLAVHL---EAEKNKHKEIT--VMVEKHK 613

Query: 70 VEFSSKQCHLRQLQQLKKLLVLQOELEFHEELOTSSYSLRQYQSILEKQTSDLVLLH 129
E S K H +Q +KL VL+Q+ + E+L+ Q + L K +++
Sbjct: 614 TELESK--H--QQDALWTEKLQVLKQQYQTEMEKREK---CEQKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHENTGKEL---HLAQQLALAGDKIASLERSLNLRYD 183
H +E + + + +E+ + + E L H +E+L+ + D+ + + L D
Sbjct: 667 HIEEMNEKTEKLDVKQTELESLSSESEVLKARHKEEELSVLKQDTDKMKELEAKMD 726

Query: 184 K---YQSSLSNIELLECQVKMLQGE--LGGIMGQEPENKGDHVKVRIYTPCMIEHQE 237
+ +Q + +I + E +V + + E L + Q + K + + + +
Sbjct: 727 EQKNHQQQVDSI--IKEHEVSIQTEKALKDQINQLELLKRDK--HLKEHQAHVENLEA 784

Query: 238 TOKRLSEVMQKVSQQDLIELRNKLCASNALVLEREKALIKLQADFASCTATHRYPPSS 297
KR Q+ S + D+ Q ++ ++ E+ L+Q T R
Sbjct: 785 DIKREGEQQASAKLDVFSYQS---ATHEQTKAYEEQLAQQLKLDLE--TERIL--- 837

Query: 298 SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRNIMKMMKL-ELD 356
+ K + ++ QK C ++ ++ V+DL +E + + +K + + E
Sbjct: 838 -----LTKQVAEVEAQKDVCTELDAHKIQVQDLMOQLEKQNSEMEQVKSITQVYESK 891

Query: 357 LH-GLREETSABIERKDKDITILQCRRL-QELQLEFETQKTLKKDKF--LQEKDEM-LQ 411
L G +E+ +K+ ILQ R Q+ ++ TOKL+ K+D L E+ E +
Sbjct: 892 LEDGNKEQEQTQKQILVEKENMILQMRQKKEIEL--TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKILTQVQNSLLK----KEKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQ 466
EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q
Sbjct: 951 NQEKMEKVKQKAKEMQETLKKLLDQEAELKKELENTALELSQ--KEKQFNAMLEMAQ 1008

Query: 467 KLNLSLEAQKQERLAAQQAQCKEEAALAGCHLEDTQRKLGKLLLDKQKADTIQELQR 526
+ +A RL Q Q + +A L D +K L Q+A+ +QE+
Sbjct: 1009 ANSAGISDAVS--RLETNQEQIESLTVHRRELNDVISIWEKKL---NQAEELQEIH- 1062

Query: 527 ELQMLQESSMAEKEQT-----SNRKR--EELSLESEALRKLNSDKEKRLQ 574
E+Q+ +KE +AE +Q K + +E ++ L +L+ K+K
Sbjct: 1063 EIQLQEKEQVAVELKQKILLFGCKEEMNEKITWLKEGVKQDTTLNELQEQKQSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHOREQGSIRCKLEEDLQEAATKLLLEDKREQLKSKSEKLM 634
++A+ + K+ L+++ + L+E L E L E+ + + + K +
Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLQEQVLKMLAEEDKRRKVELTSLKLTDD 1182

Query: 635 GELEALRQEFKKKDKRTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694
E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +
Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLFKKLEELAIQDLICCKRTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLQAQLDKALQKEKHLYLTTITKEAYDALSRKSAACQDDLT---QALE 750
K A+ + Q + K KE ++T E +A R+ Q+ L QA
Sbjct: 1242 KTNAILSR-ISHCQHRRTTK--KEALLIKTCTVSEL-EAQLRQLTEQNTLNISFQQATH 1297

Query: 751 KLNHVTSETKSLQOSLTQTEKKAQLEEEIIAYEERMKKLN---TELK--LRGFHQESE 805
+L ++ KS++ + +K L++E ++ + T+L+K + +
Sbjct: 1298 QLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEKESCITQLKELSENINAVTL 1357

Query: 806 LEVHAFDKKLE--EMSCQVLQWQKHONDKMLAAEEQLREFQEEAALKENLLEDDKE 863
++ +KK+E +S Q+ Q QN + L+ KE + +++ K LL D +
Sbjct: 1358 MKEELKKEKKEISSLSKQLTDLNVQLQNSIS--LSEKAAISSLRKQYDEEKCELL-DQVQ 1415

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGQNLRE---QVNYIAKLSG 920
++ K+ D +W K+ + + N ++E Q+ +K +
Sbjct: 1416 DLSFRVDTLSKEKISALEQVDDWSNKFSEMKKKAQSRFTQHONTKVELQIQLELKSKEY 1475

Query: 921 EKDH-LHSVMMVHLQENKK---LKKEIEEKKMAE 951
EKD ++ + L Q+NK+ LK E+E+ K K E
Sbjct: 1476 EKDEQINLLKEELDQONKRFDCLEKGMEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25
Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQKRTLQDNQLCMEEAMNS---SHD 56
 Sbjct: 470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLTREREFEQMKVALEKSQSEYLKISQE 528

Query: 57 KKQQAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQS 116
 Sbjct: 529 KEQQESLALAELELQ---KKAILTESEN---KLRLQQAETRYRTRILESSLEKSLQ 581

Query: 117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE----EMGNHNENT--GEKLHLAQEQLALA 167
 Sbjct: 582 ENKNQSKDLAVHLEAKNKHKEITVMVEKHKTELESLEKHQDQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNLYRDK---YQSSLS--NIELEECQVKMLQGEELGGIMGQEPENKGDH 220
 Sbjct: 642 MEKLREKCEQEKETLLDKKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCMIEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKL 280
 Sbjct: 701 HKLEELS--VLKD--QTDMMKQLEAKMDEQKNHQQQVDSIIEHEVSIQRTKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338
 Sbjct: 757 QINQLELLKERDKHLKEHQAHVENLEADIKRSEGEQQASAKLDVFSQYSATHEQTKA 816

Query: 339 VSEQKRNMKDMKLELDLHGLREETSABIERKOKDITILQCRLOELQLETFETQKLT 398
 Sbjct: 817 YEEQLAQLOQKLLDLETERILLTKQV-AEVEAQKDV---CT--ELDAHKIQVQDLMOQ 869

Query: 399 KDKFLQEKDEMLQLEKKTQVQNSLLKK-EKELEKQOCMATELEMTVKEAKQDKSKEAE 457
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGKKEQEQTKILVEKENMILQMRGQKKIE 925

Query: 458 C--KALQAEVQKLKNSLEEAKQOERLAAQAAQCKEEAALAGCHLEDTRK--LQKGLL 513
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFNQKKEKMEKVKQKAK---EMQETLKKKLLDQEA 981

Query: 514 DKQKADTIQEL-QRELQMLQKSSMAEKQTSNRKRVEELSLELSEALRKLNSDKERKQ 572
 Sbjct: 982 KKELENTALELSQKQKFNQAKMLEMAQANSAGISDAVSRLETNQKEQIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDKMMDMLDRIKHQHREQSGSIKCKLEEDLQEA TKLEEDKREQLKKS---KE 628
 Sbjct: 1040 LNDVISIWEKKLNQQAELQEIHEIQLOKEQEVAELKQKILLFGCEKEEMNEKITWLKE 1099

Query: 629 HEKMEGELEALRQEFKKDKTKLNSRKLEENENLRAELQCCSTOLESSLNKYNTSQ 688
 Sbjct: 1100 EGVKQDTTLNELQEQKLSAHV--NS--LAQDETKLKAHLEKLEVDLNSLKENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLQAQL---DKALQ--KEKHYLTITKEA---YDALSRSKAA 740
 Sbjct: 1156 QLVELKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLEFKKLSEE-LA 1214

Query: 741 QDDDL----TQAL-----EKLNHVTSETKSLQOSLTQTOEKKAQLEEEI IAYEERMKKL 790
 Sbjct: 1215 IQLDICCKTEALLEAKTNELINISSKTNAILSRISHCQHRRTTKVKEALLIKTCTVSEL 1274

Query: 791 NTELRLRGFQHESELEVHAFDKLEEMSCQVLQWQKHQNDLKMMLAAKEEQLEFQEM 850
 Sbjct: 1275 EAQLRQLTEEQNTLNISFQQAHOEEKENQI---KSNKADIESLVTEKEAL---QKEG 1327

Query: 851 AALKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNOLE 910
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE---NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKLLKEIEEKKMAE 951
 Sbjct: 1379 LNVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFEESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQSIL 118
 Sbjct: 165 QMLQREKKLQGISQSDKSLRRIAELEELQMDQAKHLQEEFQASLEEKQYISVL 224

Query: 119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT---GEKL---HLAQEQLALA 167
 Sbjct: 225 QTVQSLKQRLRNGPMNVVLPQLEPQAEVFTKEENPESDGEPVVDGTSVKLTETL 284

Query: 168 GDKIASLERSLNLYRDKYQSSLSNIELEECQVKMLQGEELGGIMGQEPENKGDHDKSVRIYT 227

Sbjct: 285 ++ E L ++ QS LL ++ LQ +L + QE E D ++ 340
 Query: 228 SPCMIQEHQETQKRLSEVWQVSOQDDLIQELRNKLACSNAVLEREKALIKLQADFASC 287
 +I + ++ + ++ Q +I E ++ L ++ E+ ++L++
 Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKRQM--HETLEMKEE-IAQLRSRIKQM 394
 Query: 288 TATH---RYPSSSEEC--EDIKKILKHLQEQKDSQCLHVEEYQNLVKDL-----RVE 335
 T R SE E+++K L Q+ +++ E +K + R+
 Sbjct: 395 TTQGEELREQEKESERAAFEELKALSTAQKTEARRKLAEMDEQIKTIEKTSEERIS 454
 Query: 336 LEA-VSEQKRNIMKMMKL--ELDLHLREETS AHIERKDKDITILQCRQLQELFTET 392
 L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E
 Sbjct: 455 LQQELSRVQEVV-DVMKKSSEEQIAKLKHEKELARKEQELTK---KLQTREREFEQ 510
 Query: 393 QKLTLLKKDKFLQEKDEMLOLEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAKQDK 452
 K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+
 Sbjct: 511 MKVALESQ--SEYLIKISQEKQ-----QESLAELELELQKKAIL-TESENKLRDLQGE- 561
 Query: 453 SKEAECKALQAEVQKLNKNSLEAKQER-----LAAQQAQCKEEAALAGCHLEDTOR-K 506
 ++ + L+ E L+ SL+E K Q + L A++ KE + H + + K
 Sbjct: 562 AETYRTRILELE--SSLEKSLQENKQSKDLAVHLEAKNKNKEITVMVEKHKHELESK 620
 Query: 507 LQGLLLDKQKADTIQELQELQMLQKESMAEKEQTSNRKRVEELSLELSEALRK-LEN 565
 Q+ L ++ Q+ Q E++ L +E EKE K + + E K LE
 Sbjct: 621 HQQDALWTEKLQVLKQYQTEMEKL-REKCEQEKETLLKDKKII-FOAHIEEMNEKTEK 678
 Query: 566 SDKEKRLQKTVAEQDMKNMMLDRIKHQHQGSI-KCKLEEDLQEA-TKLEDKR--E 621
 D ++ +L+ +E +++L + +H+ E+ S+ K + ++ QE K+ E K +
 Sbjct: 679 LDVKQTELESLSSE---LSEVL-KARHKLEELSULKDQTDKMKQLEAKMDEQKNHQ 733
 Query: 622 QLKKS--KEHEKLMGELEALRQEFKKDKDTLKENSRLKEEN---ENLRAELQCCSTQL 676
 Q S KEHE ++ +AL+ + + + LKE + L+E ENL A+++ +L
 Sbjct: 734 QQVDSIIKEHEVSIQRTKALKDQINQLELLKRDKHLKHEQAHVENLEADIKRSEGE 793
 Query: 677 ESSLNKYNTSQVQIQLNKEIALQKESLMSLQALQKALQKQKHLYQTITITKEAYDALSR 736
 + + K + Q +++ +E L LQ +L L+ E+ L TK+ + ++
 Sbjct: 794 QASAKLDVFSYQSATHEQTKAYEEQLAQLQKLL-LDLETERILL---TKQVAEVEAQ 848
 Query: 737 KSAACQD-----DLTQALEKLNHVTSETKSLQSLTQTEKKAQ--LEEEIIAYEE 785
 K C + DL Q LEK N SE + +SLTQ E K + +E+ +
 Sbjct: 849 KKDVTCLDAHKIQVQDLMOQLEKQ---SEMEQKVSLTQVYESKLEDGKKEQEQTKQI 905
 Query: 786 RMKMLNTELRLKRGFHOESELEVHAFDKKLEEMSCQVL--QWQKQHQNDLMLAAKEEQ 843
 ++K N L+ G Q+ E+E+ +E S +L +++ +N K + ++
 Sbjct: 906 LVEKENMILQMRG--QKKEITLQKLSAKEDSIHILNEEYETKFNQEKRMKVKQKA 963
 Query: 844 REFOEMAALKENLLEDDKEPCLPQWVPRKDTCLRYRGNDQIMTNLEQWAKQKV---- 899
 +E QE LK+ LL+ ++ L + + L + Q ++ +A+
 Sbjct: 964 KEMQE---TLKKKLLDQEA---LKK-ELENTALELSQKEKQFNKMLEQAQNSAGISD 1016
 Query: 900 ANEKLGNLREQVNYIAKLSG-EKDHLHSMVMH-LQENKKLKK--EIEEKKMAENTRL 955
 A +L +EO+ + ++ E + + S+ L Q+ ++L++ E+ + + E L
 Sbjct: 1017 AVSRLETNKEQIESLTVHRRRLNDVISIWEKKLNQAEELQEIHEIQLEKEQEVAEL 1076
 Query: 956 CTKALGPSRTESTQREKVCGLGWKGLPQD 985
 K L E + K L +G+ QD
 Sbjct: 1077 KQKIL-LFGCEKEEMNKETWLKEEGVKQD 1105
 Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25
 Identities = 220/907 (24%), Positives = 444/907 (48%)
 Query: 67 ESEVEFGSSKQCHLRQLQQLKKLLVLQLEFHTTEELQTSYSSLRQYOSILE---KQTS 123
 E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+
 Sbjct: 123 EAEDLVGNSDSLNEQLIQLRRMERSLSSYRGKYSSELVTAYQMLQREKKKLQGLSLSQ 182
 Query: 124 DLVLLHHCKLKEDEVILYEEEMGNHNTENGKHLAQEQALAGDKIASLERSLNLYRD 183
 D L +L+E+ + +++ H+ E+ +E+ I L+ +L +
 Sbjct: 183 DKSL-RRIAELREE--LQMDQAKKHLQ---EEFDASLEE---KDQYISVLQTVSLLKQ 233
 Query: 184 KYQSSLSNIELLECQVKMLQGEELGGIMGOE-PENKG-----DHKVR-ITYSPCMIOEHQ 236
 ++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +
 Sbjct: 234 RLANGPMNVDLK-PLPQLEPQAEVFTKEENPESDGEFVVEDGTSVKTLETQORVKRQE 292
 Query: 237 ETQKRLSEVWQVSOQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHRYPPS 296
 KR E Q +Q L+ KA L ER + L K+ + D T
 Sbjct: 293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTCLIT-- 346
 Query: 297 SSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKMMKLELD 356
 + D K +++ L++ K + E + + + L ++ E ++ Q R+ +K M +
 Sbjct: 347 ---QLRDAKNLIEQLEQDKGM---VIAETKRQMHEHTEMKEEIA-QLRSRIKQMTTQGE 400

Query: 357 LHGLREETS-AHIERKDKDITILQCRLOE----LQLEFTETQKLTLLKKDKFLQEKDEMLQ 411
L +E++ A E +K ++ Q + +E L+ E E K T++K +E+ + Q
Sbjct: 401 LREQEKESERAAFELEKALSTAQ-KTEEARRLKAEMDEQIK-TIEKTSE-EERISLQQ 457

Query: 412 ELEKKLTQVQNSLLKK-EKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQRLKN 470
EL + +V + + K E+++ K Q + E E+ KE Q+ +K+ + + + Q +K
Sbjct: 458 ELSRVKQEVVDVHMKSSEEQIAKLQKLH-EKELARKE--QELTKKLQTREREFQEQ-MKV 513

Query: 471 SLEAAQQERLAAQAAQCKEAAAGCHLEDQTKLQ-KGLLLD-KQKADTIQELQREL 528
+LE++ Q E L Q + +E AL L+ + + L D +Q+A+T + EL
Sbjct: 514 ALEKS-QSEYLIKISQEKQESLAELELQKKAILTESENKRLQLOQAEETYRTRILEL 572

Query: 529 QMLQKESMAEKEQTSNRKRVEELSLELSEALAKLENS-DKEKRLQKTVAEQDMKMDM 587
+ E S+ E + S V L E ++ +++ +R K +L+ +QD +
Sbjct: 573 ES-SLEKSLQENQSKDLAVH-LEAEKNKHKEITVMVEKHKTELESLKHQDALWTEK 630

Query: 588 LDRIKHQHR-EOGSIKCKLEEDLQEAATKLEDKRE--QLKKSKEHEKLMEGEALRQEF 644
L +K Q++ E ++ K E OE LL+DK Q + +EK +E +L+ + E
Sbjct: 631 LQVLKQQYQTEMEKLRKCE---QEKETLLKDKKEIIFQAHIEEMNEKTE-LKLDVVKTEL 686

Query: 645 KKKDKTLKE--NSR-KLEENENLRAELQCCSTQLESSLNKY-NTSQVQIDLNKE--IA 698
+ L E +R KLEEE L+ + +LE+ +++ N QQ + + KE ++
Sbjct: 687 ESLSELSEVLKARHKLEELSVLKDOTDKMKQLEAKMDEQKNHQQQVDSIIEKEVS 746

Query: 699 LQK-ESLMSLQA-QLDKAL-QKEKHLYLTITTEAYDALSRSK-----AACQDDLTQAL 749
+Q+ E + Q QL+ L +++KH + E +A ++S A+ + D+ Q+
Sbjct: 747 IQRTEKALKDQINQLELLKDKLKEHQAHVENLEADIKRSEGELOQASAKLDVFSY 806

Query: 750 EKLNHVTSETKSLQOSLTQTEKKAQLEEEIIAYEERMKKNTELRLKRGFQHESELEVH 809
+ H +TK+ ++ L Q Q+K LE E I +++ + + + + +
Sbjct: 807 QSATH--EQTKAYEEQLAQQLKLLDLETERILLTKQVAEVEAKQKDVCTELDAHKIQV 864

Query: 810 AFDKKEEMSCQVLQWQKHQN--DLKMLAAKEQLREFQEEMAALKENLL----EDDK 863
++LE+ ++ Q K + K+ +EQ E +++ KEN++ E K+
Sbjct: 865 DLMQQLKQNSMEQVKSLTQVYESKLEDGNKEQ--EQTKQILVEKENHILQMRGQKK 922

Query: 864 PC-CLPO-WSVPKDTCLRYRGNDQIMTNLE-QWAKQKVANE--KLGNQLREQV-NYIAK 917
L Q S +D+ + N++ T + Q K +KV + ++ L++++ + AK
Sbjct: 923 EIEILTQKLSAKEDSIHIL--NEEYETKFNQEKMEKVKQKAKEMQETLKKKLLOQEA 980

Query: 918 LSGEKDHLHSMVHLQEQENKLLKKEIEKKMKKAENTRLCTKALGPSRTSTESTOREKV 973
L K L + + L Q+ K+ ++ E M N+ + A+ SR E+ Q+E++
Sbjct: 981 L---KKELENTALELSQEKQFNAKMLE--MAQANSAGISDAV--SRLETNQEQI 1029

Score = 318 (47.7 bits), Expect = 4.4e-24, P = 4.4e-24
Identities = 184/827 (22%), Positives = 405/827 (48%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDINKLHDVCKRQRKTLQDNQLCMEAMNSHDKK-Q 59
++ E G + + S S + L+ ++ + ++ L++ ++ + D Q
Sbjct: 1323 LQKEGGNQQAASEKESCITQLKKELSENINAVTLMKEELKEKKVEISSKQLTDLNVQ 1382

Query: 60 AQ-ALAFESESEVEFGSSKQCHLRQLQOLKLLVLQOELEFHTTELQTSYYS-LRQYQS- 116
Q +++ E E S + +Q + K +LL Q+L F + L S L Q
Sbjct: 1383 LQNSISLSEKAAISSLR----KQYDEKCELQDVQDLSFKVDLSKEKISALEQVDDW 1438

Query: 117 ---ILE-KQTSIDLVLHHHCKLKEDEVILYEEEMGNHNTGKHLHLAQEQALAGDKIA 172
E K+ + H +KE ++ L + + ++ E++L +E+L +
Sbjct: 1439 SNKFSEWKKKAQSRFTQHTVKEQLQLELSKEAYEKD--EQINLLKEELDQKNKRPD 1496

Query: 173 SLERSLNLYRDKYQSSLSNIEL-LECOVKMLQELGGIMQEP-ENKGDHKSRIYTSPC 230
L+ + + K + SN+E L+ Q + EL + Q+ E + + ++ Y
Sbjct: 1497 CLKGEMEDDKSKMEKESNLETLSQTARIM-ELEDHITQKTIEIESLNEVLKNYNQK 1555

Query: 231 MIQEHQETOKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTAT 290
I EH+E ++L + ++D+ +E K+ L LE + +K + +
Sbjct: 1556 DI-EHKELVQKLQHFQELGEEDKNRVKEAEKI-----LTLENQVYSMAELETKKKELE 1609

Query: 291 HRYPPSSSEECIDIKKILKHLQEQKDSQCLHVE-EYQNLVKDLRVELEAVSEQKRNIMKD 349
H S+E E+K + L+ + ++ ++ + + ++ L+ + E+K ++
Sbjct: 1610 HVNLSVKSE-EELKALEDRLSESAKLAELKRKAQKIAIAIKKQLLSQMEEK---EE 1664

Query: 350 MMKLELDLHGLREETS-AHIERKDKDITILQCRLOELQLEFTETQKL--TLKKDKFLQEKD 407
K + H E N T + ++ ++ ++ IL+ +L+ ++ +ET + + K E++
Sbjct: 1665 QYKKGTESH--LSELNTKLQEREREVHILEEKLKSVESQSETLIVPRSAKNVAAYTEQ 1722

Query: 408 EM-----LQEL-EKKLTQVQNSLLKKEKEL-----EKQCCMATELEMTVK-EAKQDKSKE 455
E +Q+ E+K++ +Q +L +KEK L EK++ +++ EM + + + K +
Sbjct: 1723 EADSQGCVCQTYEEKISVLQRNLTEKEKLLQVQGEKEETVSSHFMCRQYQERLIKLEH 1782

Query: 456 AECKAL--QAEVQKLKNSLEAAQERLAAQAAQCK--EAAALAGCHLEDQTKLQKGL 511

Sbjct: 1783 AE K Q+ + L+ LEE ++ L Q + + + A +LE+ +QK L
 1842 AEAKQHEDQSMIGHLQEELEEKNKYSLIVAHVEEGGKNNIQAQNLNVFDDVQKTL 1842
 Query: 512 LLDKQKADTIQELQRELMQLQKESMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569
 ++K T Q L++++ L +S + +++ +R +EEL+ E +AL++++ +K
 Sbjct: 1843 ---QEKELTCQILEQKIKEL--DSCLVRQKEV-HRVEMEELTSKYEKQLAQMDGRNKP 1896
 1896
 Query: 570 KRQLQKTVAEQD---MKMNDMLDRIKHQHQREQGSICKLEEDLQEA TKLLEDKREQLKK- 625
 L++ E+ + +L ++ QH + E + Q+ K + ++ L+
 Sbjct: 1897 TELLEENTEESKSHLVQPKLLSNMEAQHNDFKLAGAEREKQKLGKEIVRLQKDLRML 1956
 1956
 Query: 626 SKEHEKLMEGEALRQEFKKKDKTLKENSRLKEENENLRAELQCCSTQLESSLNKYNT 685
 KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT
 Sbjct: 1957 RKEHQ---ELEILKKEYDQ-----EREKIKQEQEDL--ELKHNST-LKQLMREFNT 2003
 2003
 Query: 686 S-QQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLTITITKEAYDALSRSKSAACQDQ 744
 Q Q+L I ++A+L ++ Q+E + L I E D L R +A ++
 Sbjct: 2004 QLAQKEQELMENTIKETINKAQEVEAEELSHQETNQLLKXIA-EKDDDLKR-TAKRYEE 2061
 2061
 Query: 745 LTOALEKLNHVTSETSKLQOOSLTQOEKKAQ-LEEIIAYEERMK--KLNTLELRKLRGFH 801
 A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +
 Sbjct: 2062 ILDAREE--EMTAKVRDLQTOLEELQKYYQKLEQENPGNDNVTIMELQTOAQKTTLI 2119
 2119
 Query: 802 QESELEVHAFDKLEEMSCQVQWQK 827
 +S+L+ F +++ + ++ ++K
 Sbjct: 2120 SDSKLKEQEFREQIHNLEDRKKYK 2145
 2145
 Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24
 Identities = 213/977 (21%), Positives = 454/977 (46%)
 Query: 4 EAGERD-REVSSLSKLLSLQLD-IKNLHDVCKRQRTLQDNQLCMEEAMNSSHDKKQK 61
 E R+ +V S+ K L+ Q + ++ +H++ + Q K + +L + + ++ +
 Sbjct: 1034 EVHRRELNDVISIWEKKLNQQAEELEQEIHEI-LOEKEQEVAELKQKILLFGCEKEEMNK 1092
 1092
 Query: 62 ALAFESEVEFGSSKQCHLRQLQ-QLKKLL---VLOQE--LEFTEELQTSYYSLRQY 114
 + + + E G + L +LQ QLK+K + Q E L+ H E+L+ +
 Sbjct: 1093 EITLKKE---GVKQOTTLNELQELQKQSAHVNSLAQDETCLKAHLEKLEVDLNSLKE 1149
 1149
 Query: 115 QSILEKQTSDLVLLHHCKLKEDEV---ILYEEEMGNHNTGKHLHAEQQLALAGDKI 171
 + L++Q +L +L K R E+ + +E + ++ EK + + E +L K+
 Sbjct: 1150 NTFLEQLVLEKMLAEEDKRRVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLEFKKL 1209
 1209
 Query: 172 AS-LERSNLNRYDKYQSSLS--NIELLECQVMLQGLGGMQEPENKGDHSHKRVITYS 228
 + L L++ K +L EL+ L I +++ K +
 Sbjct: 1210 SEELAIQLDICCKTEALLEAKTNELINISSKTNAILSRI--SHCQHRTTKVKEALLK 1267
 1267
 Query: 229 PCMIQEHQ-----ETQKRLSEVMQKVSQ-ODLIQELRNKLACSNALVLEREALIKL 280
 C + E + E Q L+ +Q+ + Q ++ +++ A +LV E+E L
 Sbjct: 1268 TCTVSELAQLRQLTEQNTLNISFQATHQLEEKENQIKSMKADIESLVTEKA---L 1323
 1323
 Query: 281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340
 Q + + + S E C I ++ K L E ++ L EE +K+ +VE+ ++S
 Sbjct: 1324 QKEGGN---QQQAASEKESC--ITQLKKESENINAVTLMKEE---LKEKKVEISSLS 1373
 1373
 Query: 341 EQKRNIMKDMKLELDLHGLREETSABIERKDKDITILQCRQLQEL--QLEFTETQKLT-L 397
 +Q ++ + + L S+ ++ D++ L ++Q+L +++ +K++ L
 Sbjct: 1374 KQLTDLNVQLQN-SISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKVDTLSKERISAL 1432
 1432
 Query: 398 KK-DKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQOQMATELEMTV---KEAKQDKS 453
 ++ D + + E ++ + + TQ QN++ + + +LE + A E + + KE ++
 Sbjct: 1433 EQVDDWSNKFSEWKKKAQSRFTQHNTVKELQIQLELKSKEAYEKDEQINLLKEELDQQN 1492
 1492
 Query: 454 KEAECKALQAEVQKLKNSLEEAQQRERLAAQQAQCKEEAALAGCHLE-DTQKRLQGLL 512
 K +C + E K K +E+ + L +Q A + E + +E ++ ++ K
 Sbjct: 1493 KRFDCLKGEMEDDKSMKKEKSNLETELKSTARIMELEDHITQXTIEIESLNEVLKNY- 1551
 1551
 Query: 513 LDKQKADTIQELQRELMQLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRO 572
 + +QK +EL ++LQ Q+ + +++ L ++ +LE KE
 Sbjct: 1552 -NQKNDIEHKELVQKLQHFQELGEEKDNRVKEAEKILTLENQVYSMAELETKKKELEH 1610
 1610
 Query: 573 LQKTVAEQDMKNDMLDRIKHQHQREQ-GSICKLEEDLQEA TKL---EDKREQLKSK 627
 + +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK
 Sbjct: 1611 VNLVSKSEELKALEDRLESESAKLAELKRKAQKIAIKQLLSQMEKEEQYKKG 1670
 1670
 Query: 628 EHEKLMEGEALRQEFKKKDKTLKENSRLKEE-ENENL---RAELQCCSTQLESSLNK 682
 E EL OE +++ L+E + +E +E L A+ T+ E + ++
 Sbjct: 1671 ESHL---SELNTKQEREREVHILEEKLKSVSSQSETLIVPRSAKNVAAYTEQEEADSQ 1727
 1727
 Query: 683 ---YNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLTITITKEAYDALSRSKA 739
 T ++ I L + + +KE L+ Q +K H+ +E L A
 Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFMRCQYQERLIKLEHAEA 1785
 1785

Query: 740 ACQDDLTQALEKLNHVTSET--KSLQSSLTQTQEKKAQLEEEIIAYEERMKKLNTELRLK 797
+D Q++ + H+ E K+ + SL Q + + + I ++ + + + ++K
Sbjct: 1786 KQHED--QSM--IGHLQEELEENKKYSLIVAQHVEKEGGKNNIQAQNLNENFDDVQRT 1841

Query: 798 RGFHQSELEVHAFDCKKLEEM-SCQVLQWQKQHQNDLKMALAAKEEQREFQEEMAALKEN 856
QE EL ++K++E+ SC V Q ++ H+ +++ L +K E+L+ Q+ K
Sbjct: 1842 L---QEKELTCQILEQKIKELDSCLVRQ-KEVHRVEMEELTSKYEKLOALQMDGRNKPT 1897

Query: 857 -LLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQKVANKEKLNQLEQVNYI 915
LLE++ E PK + ++ + L A+++ +KLG ++ +
Sbjct: 1898 ELLEENTEESKSHLVQPKLLSNMEAQHNDLEFKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query: 916 AKLSGE-KDHLHSMVHLQGENK-KLKKEIEKKMKAENTRLCTKALGPSRTESTOREK 972
L E + L + QE + K+K+E E+ ++K +T + + T+ Q+E+
Sbjct: 1954 RMLAKEHQEELILKKEYDQEREKIKQEEDLELKHNS--LKQLMREFNTQLAQKEQ 2010

Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22
Identities = 221/952 (23%), Positives = 441/952 (46%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDINKLHDVCRQRKTLDQNL---CMEEAMNSSHD- 56
+K A E R+VS L SKL + + ++L ++ K+L+D L + E + D
Sbjct: 1160 LKMLAEEDKRKVSLETSKLTDEEFQSLKSSEKSNKSLDSKLEFKLSEELAIQLDI 1219

Query: 57 --KKQAQALAFEESE-VEFGSSK-QCHLRQLQOLKKLLVLQOELEFHT---EELQTSYY 109
KK L + +E + SSK L ++ + + ++ L T EL+
Sbjct: 1220 CCKKTEALLEAKTNELINISSKTNAILSRISHCQHRRTTKVKEALLIKTCTVSELEAQLR 1279

Query: 110 SLRQYQSIKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNTGKELHLAQE---QLAL 166
L + Q+ L H + KE+++ ++ EK L +E Q
Sbjct: 1280 QLTEEQNTLNISFQAT---HQLEKENQIKSMKADI---ESLVTEKEALQKEGGNQQA 1333

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQELGGIMGQEPENKGDHKSURIY 226
A +K E + + + + L+ ++K + E + Q + V++
Sbjct: 1334 ASEK---ESCITQLKKESENINAVTLMKEELKEKKEISSLKQLTD----LNVQLQ 1384

Query: 227 TSPCHIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKLQADFAS 286
S + ++ + ++ + D +Q+L K+ + L E+ AL ++ D+++
Sbjct: 1385 NSISLSEKEAAISSLRKYDEEKCELLDQVQDLSFRV---DTLSKEKISALEQVD-DWSN 1440

Query: 287 CTATHRYPPSS--SECEIDIKILKHQEQKDSQCLHVEEYQNLVKD-----LRVE-LE 337
+ S + + +K++ L E K + +E NL+K+ R + L+
Sbjct: 1441 KFSEWKKKAQSRFTQHQNTVKELQIQIQL-ELKSKEAYEKDEQINLLKEELDQKQNRFCCLK 1499

Query: 338 AVSEQRNIM-KDMMKLELDLHGLRE---ETSAHIERKDKDITILQCRQLQEL-QLEFET 392
E ++ M K LE +L E HI +K +I L L+ Q + E
Sbjct: 1500 GEMEDDKSKMEKKESNLETELKSTARIMEEDHITQKTIIESLNEVLKNYNQKQDIEH 1559

Query: 393 QKLTLLKDKFLQ---EKDEMLQELEKLLTQVQNSLLKKEKELEKQOCMATELEMTVKEAK 449
++L K F + EKD ++E E+K+ ++N + + ELE ++ + ++VK
Sbjct: 1560 KELVQKLQHFQELGEEKDNVRKEAEKILTLENQVYSMKAELETKKKELEHVNLSVK--- 1616

Query: 450 QDKSKEAECKALQAEVQKLNKLSLEAKQOERLAAQQAQCKEEAALAGCHLEDTQRKLQK 509
SKE E KAL+ ++ S + + +R A Q+ A K++ +E+ + +K
Sbjct: 1617 ---SKEELKALEDRLES--ESAARKLAELKRKAQKIAAIKKQLL---SQMEEKEEQYKK 1668

Query: 510 GLLLDKQKADT-IQELQRELQMLQKESMAEKEQTSNRRKRVESLSLEALRKLNSDK 568
G + +T +QE +RE+ +L+++ E Q+ + S + A + E +D
Sbjct: 1669 GTESHLSELNKLQEREREVHILEEKLSVESSQSETL--IVPRSAKNVAAYTEQEEADS 1726

Query: 569 E---KRQLQK-TVAEQDMKMNO-MLDRIKHOREGGSICKLEEDLQEAATKLEDDKREQ 622
+ K +K +V +++ + +L R+ Q +E+ ++ E Q +L+ K E
Sbjct: 1727 QGCVRQTYEEKISVLORNTEKEKLLQVRG-QEKEE-TVSSHFMRCQYQERLI--KLEH 1782

Query: 623 LKKSKEHE-KLMGEL-EALRQEFKKDKTLKENSRLKEENENLRAELQCCSTQLESSL 680
+ +K+HE + M G L E L +K K + +K E N+++ +L
Sbjct: 1783 AE-AKHEDQSMIGHLQEELEENKKYSLIVAQHVEK-EGGKNNIQAQ-----QNLE--- 1832

Query: 681 NKYNTSQQVIODLNKEIALQKESLMSLAQLDKAL--QKEKHYLTITKEAYDALSR-K 737
N ++ Q+ +Q+ KE+ Q L +LD L OKE H ++ Y+ L +
Sbjct: 1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVRQKEVHRVEMEELTSKYEKLOALQ 1888

Query: 738 SAACQDDLTQALEKLNHVTSETSKSLOSLTQTEKKAQ-LEEEIIAYEERMKKLNTEL-- 794
++ T+ LE+ S++ +Q L E + LE ++ E +KL E+
Sbjct: 1889 QMDGRNKPTELLEENTEESKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVR 1948

Query: 795 --RKLGRGFHQSELEVHAFDCKKLEEMSCQVLQWQKQHQNDLKMALAAKEEQREFQEEMA 852
+ LR +E + E+ K+ ++ + ++ Q+Q +LK + + +REF ++A
Sbjct: 1949 LQKDLRMLRKEHQEELILKKEYDQEREKIK-QEEDLELKHNSLTKQLMREFNTQLAQ 2007

Query: 853 LKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQKVANKEKLNQLEQV 912

Sbjct: 2008 ++ L KE Q V + + Q TN Q K K+A EK + R
KEQELMTIKETINKAQ-EVEAEELSH---QEETN--QLLK--KIA-EKDDDLKRTAK 2057
Query: 913 NYIAKLSGKDLHLSVMVHLQOENKKLKEIEKKMKAEN 952
Y L ++ + + LQ + ++L+K+ ++K + EN
Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQKLEQEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22
Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLNKLQSLQDIKN--LHDVCKRQRKTLQDNQLCMEAMSSSHDKK 58
+KD+ + +N K L +LD+K L + +L+ +EE ++ D+
Sbjct: 657 LKDKKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEEELSVLKDQT 714
Query: 59 QAQALAFESEVEFGSSKQCHLRQLQKLLV-LQOELFHTLELQTSYSSLRQYQSI 117
+E E + K H +Q+ +K+ V +Q+ + +++ L++
Sbjct: 715 DKMK--QLEAKMDEQNNHHQQQVDSIIKEHEVSIQTEKALKDQINQLELLKERDKH 771
Query: 118 LEKQTSDLVLLHHCKLKEDEVILYEEEMG---NHNENTGEKHLAQEQLALAGDKIASL 174
L++ + L K E E+ ++ ++ T E+ +EQLA K+ L
Sbjct: 772 LKENQAHVENLEADIKRSEGLQOASAKLDVFSYQSATHEQTKAYEQLAQQLKLDL 831
Query: 175 ERSNLNRYDKYQSSLSNIELLECQVKMLQGEGLGIMGO-EPENKGDHVKVRIYTSPCMIQ 233
E L + + + + + + +M Q E +N KV+ T
Sbjct: 832 ETERILLTKQVAEVEAQKDVCTELDAHKIQQDLMOLEKQNSEMEQKVKSLTQ-VYES 890
Query: 234 EHQETQKRLSEVWQKVSQDDLIQELRN---KLACSNAVLEREKALIKLQADFASCTA 289
+ ++ K + Q + ++++I ++R ++ + +E ++ L ++ +
Sbjct: 891 KLEDGNKEQEQTQKILVEKENMILQREGQKKEITLTKLSAKEDSIHLNEEYET--- 947
Query: 290 THRYPPSSSEECEDIKKLKLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349
++ + ++ E +K+ K +QE + L E L K+L +S++++
Sbjct: 948 --KFK-NQEKMEKVKQKAKEMQETLKKLLDQEA--KLKKELENTALELSQEKQFNAK 1002
Query: 350 MMKL-ELDLHGLAETS-A-HIERKDKDITILQCRLEQLQLEFTETQKITLKKDKFLQEKD 407
M+++ + + G+ + S +K++ ++ + +EL + +K ++ +LQE
Sbjct: 1003 MLEMAQANSAGISDAVSRLETNQKEQIESLTVHRRELNDVISIMEKLNQQAELQEIH 1062
Query: 408 EM-LQELEKKLTQVQNSLLK---KEKELEKQOCMATE---LEMTVKAQD-KSKEABC 458
E+ LQE E+++ ++L +++E+ K+ E +T+ E++ K+ K A
Sbjct: 1063 EIQLQKEQEVAELQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQELQKQSAHV 1122
Query: 459 KALQAEVQKLNLSLEAAQOERLAAQAAQCEEAALACGCHLEDTOQKLOGLLLDKQKA 518
+L + KLK LE+ + + + + +E+ +E+ +RK+ + L K K
Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLOQLVELKMLAEEDKRVSE--LTSKLKT 1180
Query: 519 OTIQELQRELQMLQKSSMAEKEQTSNRKRVEELSLESEALRKLKLENSDKRQLQKTVA 578
T+E Q +K+ E + +K EEL++L +K E +K + +
Sbjct: 1181 -TDEEFQSLKSSHEKSNSKLEDSLEFKKLSEELAIQLDICKKTEALLEAKTN--ELIN 1237
Query: 579 EQDMKMNMDLRIKH-QHREQGSICKLEEDLQEATKLEEDKREQLKKSKEHEKLMGEG 637
K N +L RI H QHR K++E L T + + QL++ E + +
Sbjct: 1238 ISSSKTNAILSRISHCQHRTT-----KVKEALLIKTCTVSELAQLRLTEEQNTLNISF 1292
Query: 638 EALRQEFKKKD---KTLKNSRKLEENENLR-----AELQCCSTQLESSL---- 680
+ + ++K+ K++K + L E E L+ +E + C TQL+ L
Sbjct: 1293 QOATHQLEEKENQIKSMKADIESLVTEKALQKEGGNQQAASEKESCITQLKKELSENI 1352
Query: 681 NKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQ-KEKHYLQTTITKEAYDALSRKSA 739
N ++ ++ EI+ +L L QL ++ EK +++ K+ YD +
Sbjct: 1353 NAVTLKMEELKEKKEISSLKQLTDLNVQLQNSISLSEKEAAISSLRKQ-YDEEKCELL 1411
Query: 740 ACQDDLTQALEKLN-HVTSETKSLQSLTQTEKKAQLEEEIIAYEERMKKLNTLR-KL 797
DL+ ++ L+ S + + + E K + + ++ +K+L +L K
Sbjct: 1412 DQVQDLSFKVDTLSEKISALEQVDDMSNKFSEWKKQASRFTQHONTVKELQIQLELKS 1471
Query: 798 RGFHQSELEVHAFDKKLEEMSCQVLQWQKHQNDKMLAAKEEQLR-EFQEEAALKEN 856
+ +++ E +++ ++L++ + + + +D + KE L E + + A + E
Sbjct: 1472 KEAYEKDE-QINLKEELDQONRRFOLKGEMEDDKSKMEKKESNLETLSQARIME- 1529
Query: 857 LLEDDKEPCCLPQWSPKDTCLRYRGNDQIMTNLEQWAKQKVANEKLGNLREQVNYIA 916
LED + + T + N+ ++ N Q OK K +L +++ +
Sbjct: 1530 -LEDH-----ITOKTIEIESLNE-VLKNYNQ---QKDIEMK---ELVQKLQHFQ 1570
Query: 917 KLSGKDH---LHSMVHLQOENKKLKEIEKKMKAENTRLCTKA 959
+L EKD+ ++ L+ + +K E+E KK + E+ L K+
Sbjct: 1571 ELGEEKDNRVKEAEKILTLENQVYSHKAELETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22
Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMNSSHDKQAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTELQT 106
 + E N + + Q EE E + S K ++ L + LQ+E +
 Sbjct: 1281 LTEEQNTLNISFQOATHQLEEKENQIKSMKA----DIESLVTEKALQKEGNNQQAASE 1336

Query: 107 SYYSLRQYQSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNTGEKHLHAEQLAL 166
 + Q + L + + + L+ K K+ E+ +++ + N + L++++ A
 Sbjct: 1337 KESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLRYDRKYSSLSNIELLECQVMKQELGGIMGQEPENKGDHSHKVR 226
 I+SL + Y ++ L ++ L +V L E + Q + S+ +
 Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVWQKVSQODDLIQEL--RNK-LACSNALVLE--- 272
 + +HQ T K L E ++K Q + L +EL +NK C + +
 Sbjct: 1448 KAQSRFTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQONKRFCDLKGEMEDDKS 1507

Query: 273 -REKALIKLOADFASCSTAT---HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQN 327
 EK L+ + S TA + + E E + ++LK+ +QXD E+ +
 Sbjct: 1508 KMEKKESNLETELKSTARIMELEDHITQKTIEIESLNEVLKNYNQXDI-----EHKE 1561

Query: 328 LVKDLRVEAEVSEQKRNIMKDMKLELDLHGLREETSABIERKDKDI--TILQCLQEL 385
 LV+ L+ + + E+K N +K+ + L L A +E K K++ L + +E
 Sbjct: 1562 LVQKLQ-HFQELGEERDNVKEAEKILTLENQVSMKAELETKKKELEHVNLSVKSKEE 1620

Query: 386 QLETTETQKLTLLKDKFLQEKDEMLOLEKKLTQVNSLLKKEKELEKQOCMATELEMTV 445
 +L+ E + L+ + + E+ ++ E+K+ ++ LL + +E E+Q TE ++
 Sbjct: 1621 ELKALEDR--LESES-AAKLAELKRAEKQIAAIAKKQLLSQMEKEEQYKKGTESHLSE 1676

Query: 446 KEAKQDKSEAECKALQAEVQKLKNSLEEAQOERLAAQQAACK-EAAALAGCHLEDTQ 504
 K + +E E L+ +++ ++S E RA AA + EEA GC + +
 Sbjct: 1677 LNTKLQE-REREVHILEEKLSVSSQSETLIVPRSAKNVAAYTEQEEADSQGCVCQKTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQELQMLQKESMAEKEQTSNRKRVESLSEALRKLE 564
 K+ +L + + + LQR Q +KE +++ + R + +E ++L A K
 Sbjct: 1736 EKIS--VLQRNLTEKELLRVQ--EKEETVSSHFM--RCQYERLILKEHAERKQ 1788

Query: 565 NSDKERKLOKTVAEQDMKMDMLDRIKHQHREQ--SIKCK--LE---EDLQ-----E 611
 LQ+ + E+ + K + + + H +E G +I+ K LE +D+Q E
 Sbjct: 1789 EDQSMIGHLQEELEENKYSLLV--AQHVEKEGGKNIAQKQLENVFDVQKTLQEKE 1846

Query: 612 AT-KLEDKREQLKKSKEHEKLMEG-ELEALRQEFKKDKTLKENS---KLEENENL 665
 T ++LE K ++L +K + E+E L ++K + R +L EEN
 Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQMDGRNKPTELLEENTEE 1906

Query: 666 RAEQQCCSTQLESSLN-KYNTSQVQIDLNKEIALQKESLMSLQAOLQKALQEKHYLOT 724
 ++ +L S++ ++N + + +E + ++ ++ LQ L + L+KE H +
 Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDL-RMLRKE-HQEL 1964

Query: 725 TITKEAYDALSRKSAACODDLTQALEKLNHVTSETKSLQOSLTQTOEKKAQLEEEIAYE 784
 I K+ YD R+ Q+ + LE L H ++ + ++ TQ +K+ +LE I +
 Sbjct: 1965 EILKKEYDQ-EREKIKQE--EDLE-LKHNSLTKQLMREFNTQLAQKEQLEMTI---K 2017

Query: 785 ERMKKLNTLRLKRGFHOSELEVHAFDKKLEEMSCQVLOWQKHQNDLKMALAAKEEQLR 844
 E + K +L HQE E + KK+ E + + K+++ ++L A+EE++
 Sbjct: 2018 ETINKAQEVEAEELLESHE--ETNQLLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEMAALKENLLEDDKEPCCLPQWVSP-KDTCRLYRGNDQIMTNLEQWAKQKQVAN 903
 ++ E L + ++ L Q P D + ++ T L Q K ++K
 Sbjct: 2072 AKVRDLQTLQELQKQYQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTTLISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDLHLSVMV-HL 932
 L Q REQ++ + +L + ++++ V HL
 Sbjct: 2124 LKEQEFREQIHNLDRLLKRYEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20
 Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQDLIKN-LHDVC-KRQRTLQDNQLCMEAM-NSSHDKK 58
 ++ ++E+ +L KLL + +K L + + +K Q N +E A NS+
 Sbjct: 957 EKVQKAKEMQETLKKLLDQEAKKLELNTALELSQKEQFNKMLEMAQANSAGISD 1016

Query: 59 QAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTELQTSYYSLRQYQSIL 118
 L + E + S + H R+L + + + ++L EELQ + ++ +
 Sbjct: 1017 AVSRLETNKE-QIESLVEVHRRLENDV---ISIWKKLNQAAEELQ-EIHEIQLQEK-- 1069

Query: 119 EKQTSDLV--LLHHHCKLKE-DEVILYEEEMGNHNTGEKHLHAEQLALAGDKIASLE 175
 E++ ++L +L C+ +E ++ I + +E G + T +L +Q + + +A E
 Sbjct: 1070 EQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQELQKSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVMKQELGGI--MGQEPENKGDHSHKVRITSPCMIO 233
 L + +K + L N L E LQ +L + + +E + K ++ T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQVLKMLAEEDKRKVELTSKLTTODEEFQ 1186

Query: 234 E----HQETQKRLSEVWQKVSQDDLIQELRNKL--AC--SNALVIEREKALIKLQADFA 285
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNKSLD---KSLEFKKLEELAIQLDICKKTEALLKAKTNELINISSKT 1243

Query: 286 SCTATH-RYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRR 344
+ + + + + + + + I + + +Q + E QN + + + E+K

Sbjct: 1244 NAILSRISHCQHRTTKVKEALLIKTCTVSELEAQLRLTEEQNTLNISFQQATHQLEKEE 1303

Query: 345 NIMKMMKLELD-LHGLREETSABIERKDKDITILQCRLEQLQLEFTE--QKLTLLKDKF 402
N +K M K +++ L +E + + + + + +L+ E +E +TL K++

Sbjct: 1304 NQIKSM-KADIESLVTEKALQKEGNGQQAAASEKSCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQKDEMLQLEKLLTQVNSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQ 462
L+EK + L K+LT + N L+ L + + + + L E K + + + L

Sbjct: 1362 LKEKKEVEISSLSKQLTDL-NVQLQNSISLSEKEAAISSLRKQYDEEKCELLDQVQ--DLS 1418

Query: 463 AEVOKLNSLEEAQOERLAAQAAQCKEEAALACGHLEDTORKLQKGLLLDKQKA---- 518
+V L L A +Q + + + + K++A + +T + +LQ L L + +A

Sbjct: 1419 FKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTOHONTVKELQIQLELKSKEAYEKD 1478

Query: 519 DTIQELQRELOMLQKESMAEKQTSNRKRVEELSLELSEALRKLENSDKERQLOKQTV 578
+ I L+ EL K + E + + + +E+ L + + + +L+ +

Sbjct: 1479 EQINLLKEELDQKNRFDCLKGEHEDDKSKMEKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMDMLDRIKHQHREQGSIKCK-LEEDLQEATKLEDKREQLKKSHEKLMEGEL 637
+ + + + + + +K+ + +Q I+ K L + LQ +L E+K + +K++E +E + +

Sbjct: 1536 QKTIEIESLNEVLKN-YNOQKDIEHKELVOKLQHFQELGEEKDNRVKEAEKILTLENQV 1594

Query: 638 EALQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLES-SLNKYNTSQVQIQDLNKE 696
+++ E + K K L+ + + + E L+A L+ +LES S K + + + +

Sbjct: 1595 YSMKAELETKKLEHVNLSVKSKEEELKA-LE---DRLESESAKL---AELKRKAQEK 1647

Query: 697 IALQKESLMSLAQQLDKALQKEKHYLQTTITKEAYDALSRSACODDLTQALEKLNHVT 756
IA K+ L+S Q++ +KE+ Y + T + + L+ K + + + EKL V

Sbjct: 1648 IAAIKQLLS---QME---EKEEQYKKG--ESHLSLNTKLQEREREVHILEEKLKSVE 1699

Query: 757 S---ET---KSLQSLTQTEKKAQLEEEII-AYEERMKKLNTLRLKRGFHQSELEV 808
S ET +S + T++++A + + YEE++ L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEADSQCCVQKTYEEKISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKKLEMSQVQLWQKHQNDLKMALAAKEEQLREFQEEMAALENLLEDDKPECCPL 868
+ + EE + + Q+Q L L L E + E Q + L+E L E +K+ +

Sbjct: 1753 QRVGQKEETVSSHFMRCQYQERLIKLEHAEAKQHEQDSMIGHQEELEENKKYSLIV 1812

Query: 869 QMSVPKDTCLRYRGNDQIMTNLEQ-WAKQOKVANER-LGNQLREQ-VNYIAKLSGEKDH 925
V K+ + N Q NLE + QK EK L Q+ EQ + + + +

Sbjct: 1813 AQHVEKEGGK---NNIAQAKNLENVFDVQKTLQEKELTCQILEQKIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQQENKKLK 940
H V M L + +KL+

Sbjct: 1870 HRVEMEELTSKYEKLO 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHQETQKRLSEVWQKVSQDDLIQE-LRNKLACSNALV-LEREKALIKL-QADFASCTA 289
+E +TQ ++ +V + L + ++ L S++ L R + + L + D S TA

Sbjct: 53 RESGDTQSFQKQLRLVPSVESLFRSPIKESLFRSSKESLVRTSSRESLNRLDLSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRRNIMKD 349
+ P E ED+ L + + + Q L + + R+ + + + + + +

Sbjct: 113 SFDPSSDMDSAEEDLVGNSDSLKKEQLIQRLR--RMERSLSSYRGKYSLEVAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSABIERKDKDIT-ILQCRLEQLQLEFTE--QKLTLLKDKFLQEKDE 408
KL+ G+ ++ +DK + I + R +ELQ++ + L + D L+EKD+

Sbjct: 171 KKKLQ---GILSQS-----QDKSLRRIAEER-EELQMDQAKKHLQEEFDASLEEKDQ 219

Query: 409 MLQLEKLLTQVQNSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQAE---V 465
+ L+ + + + + L + + + + +LE + + + + +E+ + + + + V

Sbjct: 220 YISVLQTVSLLKQLRLRNGPMNVOLKPLP-QLEPQAEVFTKEENPESDGEPPVEDGTSV 278

Query: 466 QKLNLSLEEAQOERL--AQAAQOC-KEEAALACGHLEDTORKLQKGLL-LDKQKADTI 521
+ L+ + K+QE L ++ Q KE+ L E Q +L + L L+K K +

Sbjct: 279 KTLETQQRVKQENLLKCKETIQSHKEQCTLLTSEKALQEQDLERLQLEKIKDLHM 338

Query: 522 QELQRELOMLQKESMAEKQTSNRKRVEELSLELSEALRKLENSDKERQLOKQTVAEQD 581
E + + L+ + + E+ + + E + + E L E + R K + Q

Sbjct: 339 AEKTKLITQLADAKNLIEQLEQDKGMVIAETKQRMHETLEMKEEIIAQLRSRIKQMTTQG 398

Query: 582 MKHNDMLDRIKHQHQREOGSIKCKLEEDLQEAT-KLEEDKREQLK---KSKEHEKL-MEGE 636
++ + ++ + E+ + +EA KL + EQ+K K+ E E++ ++ E
Sbjct: 399 EELREQKEKSERAAFELEKALSTAQKTEEARRLKAEMDEQIKTIEKTSEERISLQQE 458

Query: 637 LEALRQEFKK-KDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQVQIDLNK 695
L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K
Sbjct: 459 LSRVKQEVVDVMKSSSEEQIAKLQKLHEK---ELARKEQLTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLMSLQAQLDKALQKEKHYLTITITKEAYDALSRKSAACQDDLTQALEKLN-H 754
+AL+K L+ +K Q+ + + K+A S DL Q E
Sbjct: 513 -VALEKSQSEYLLKISQEKQESLALAELELQKKAILTESENKLR---DLQQEAETYRTR 568

Query: 755 VTSETKSLQQLSTQTQEKKAQLEEEIIAYEERMKKLNTELRLKRGFHESELEV--HAFD 812
+ SL++SL QE K Q ++ + E K N E+ + H+ +ELE H D
Sbjct: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAKNKHNEITVMVEKHK-TELESKHKQD 624

Query: 813 KKLEEMSCQVLQWQKHQNDLKMMLAAKEEQLE-----FOEEMAALKENLLED-DK 862
E QVL+ +Q+Q +++ L K EQ +E FO + + E LE D
Sbjct: 625 ALWTE-KLQVLK---QQYQTEMEKLEKCEKETLLKDKKEIIFQAHIEEMNEKTLEKLDV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKVANEKLGQNLREOVNIAKLSGEK 922
+ L S+ + + + ++ L Q ++L ++ EQ N+ +
Sbjct: 682 KQTELE--SLSSELSEVLKARHKLEELSVLKQDTDKMKQELEAKNDEQKNHHQQQVDSI 739

Query: 923 DHLHSMVMHLQEQENKKLKEIEEKKM 948
H V + Q+ K LK +1 + ++
Sbjct: 740 IKEHEVSI--QRTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09
Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQELEKKLTVQNSLLKKEKELEKQCCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467
M ++L++K+++ Q L + + +T M + + + E + Q
Sbjct: 1 MFKKLKQKISSEEQQLQALAPAQASSNSSTPTRMRSRTSSFTQLDEGTPNRESGDTQS 60

Query: 468 LKNSLE-EAKQQRERLAAQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA--DTIQEL 524
L+ E L + ++ + + R+ L LD A D ++
Sbjct: 61 FAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTASFDPPSDM 120

Query: 525 QRELQMLQKESMAEKEQTSNRKRVEELSL-----ELSEALRKLNSDKKRLQKTVAE 579
E + L S + EQ R R E SL + SE + + +EK++LQ +++
Sbjct: 121 DSEAEDLVGNSOSLNKEQLIQLRLRMERSLSSYRGKYSYELVTAYQMLQREKKKLQGILSQ 180

Query: 580 -QDMKNDMLDRIKHQHQREOGSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632
QD + + + + +Q + K EE L+E + +L+ + LK+ + +
Sbjct: 181 SQDKSLRIAELELQMDQAKKHLQEEFDASLEEKDQYISVLQTQVSLKQLRLNGPM 240

Query: 633 MEGEALALRO-EFKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQ 688
L+ L Q E + + T +EN E E+ L+ +++ N ++
Sbjct: 241 NVDVLKPLPQLEPAEVFTKEENPESDGPVVEDGTSVKTLETLQQRVKRQENLLKRCKE 300

Query: 689 VIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLTITITKEAYDALSRKSAACQDDLTQA 748
IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +
Sbjct: 301 TIQSHKEQCTLLTSEKALQEQLDERLQ-ELEKIKDLHMAEKTCLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQQLSTQTQEKKAQLEEEIIAYEERMKKLNTELRLKRGFHESELE 807
LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E
Sbjct: 358 LEQDKGMVIAETK---RQMHELEMK---EEEIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKHQNDLKMMLAAKEEQLEFQ-----EEMAALKENLLEDDE 863
AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E
Sbjct: 410 RAAF---EELKALSTAQKTEEARRLKAEMDEQIKTIEKTSEERISLQQELSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKVANEKLGQNLQ-----EQVNYIAK 917
+ + S + +L + +++ + EQ K+ + + Q++ Q Y+ K
Sbjct: 466 VVDVMKSSSEEQIAKLQKLHEKELARKEQLTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDHLHSMVMH-LQENKKLKEIEEK---KMKAEINTRLCTKALGPSRTESTOREK 972
+S EK+ S+ + L+ + K+ E E K + +AE R L S +S Q K
Sbjct: 525 ISQEKQESLALAELELQKKAILTESENKLRDLQQAETYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3_lg13, frame 1

Report for DKFZphtes3_lg13.1

[LENGTH] 1007
[MW] 117480.77
[PI] 5.90

[HOMOL] TREMBL:AF092090_1 product: "cpl51"; Rattus norvegicus cpl51 mRNA, partial cds.
0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[EC] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04
[PIRKW] 3.6.1.32 Myosin ATPase 1e-16
[PIRKW] nucleus 3e-10
[PIRKW] phosphotransferase 6e-09
[PIRKW] duplication 2e-06
[PIRKW] citrulline 2e-12
[PIRKW] tandem repeat 1e-16
[PIRKW] endocytosis 2e-13
[PIRKW] heart 8e-13
[PIRKW] transmembrane protein 1e-13
[PIRKW] serine/threonine-specific protein kinase 6e-09
[PIRKW] zinc finger 2e-13
[PIRKW] metal binding 2e-13
[PIRKW] DNA binding 4e-12
[PIRKW] muscle contraction 1e-16
[PIRKW] acetylated amino end 1e-11
[PIRKW] actin binding 1e-16
[PIRKW] mitosis 5e-15
[PIRKW] microtubule binding 5e-15
[PIRKW] ATP 1e-16
[PIRKW] thick filament 1e-16
[PIRKW] phosphoprotein 4e-16
[PIRKW] skeletal muscle 2e-14
[PIRKW] calcium binding 2e-12
[PIRKW] alternative splicing 1e-16
[PIRKW] coiled coil 1e-16
[PIRKW] P-loop 1e-16
[PIRKW] heptad repeat 3e-10
[PIRKW] methylated amino acid 1e-16
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 2e-13
[PIRKW] cardiac muscle 8e-13
[PIRKW] hydrolase 1e-16
[PIRKW] microtubule 3e-10
[PIRKW] muscle 8e-13
[PIRKW] EF hand 2e-12
[PIRKW] cytoskeleton 2e-15
[PIRKW] hair 2e-12
[PIRKW] calmodulin binding 2e-13
[PIRKW] Golgi apparatus 3e-10
[SUPFAM] myosin heavy chain 1e-16
[SUPFAM] conserved hypothetical P115 protein 1e-07
[SUPFAM] centromere protein E 5e-15
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09
[SUPFAM] calmodulin repeat homology 2e-12
[SUPFAM] myosin motor domain homology 1e-16
[SUPFAM] alpha-actinin actin-binding domain homology 2e-07
[SUPFAM] plectin 2e-07
[SUPFAM] trichohyalin 2e-12
[SUPFAM] pleckstrin repeat homology 8e-08
[SUPFAM] ribosomal protein S10 homology 2e-07
[SUPFAM] giantin 3e-13
[SUPFAM] protein kinase homology 6e-09
[SUPFAM] protein kinase C zinc-binding repeat homology 8e-08
[SUPFAM] kinesin motor domain homology 5e-15
[SUPFAM] human early endosome antigen 1 2e-13
[SUPFAM] M5 protein 1e-07
[PROSITE] LEUCINE_ZIPPER 7
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 20

```

{PROSITE} TYR_PHOSPHO_SITE 1
{PROSITE} PKC_PHOSPHO_SITE 16
{PROSITE} ASN_GLYCOSYLATION 2
{KW} All_Alpha
{KW} LOW_COMPLEXITY 15.00 %
{KW} COILED_COIL 42.40 %

```

```

SEQ MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNSSHKKKQA
SEG .....XXXXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ QALAFESEVEFGSSKQCHLRQLQQLKKLLVLQQLLEFHTTELTQSYSLRQYQSILEK
SEG .....XXXXXXXXXXXXXXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ QTSDLVLLHHHCKLKEDEVILYEEEMGNHNTGKHLHAQEQALAGDKIASLERSLNL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ YRDKYSSLSNIELLECQVKMLQGLGGIMGQEPENKGDHVKVRIYTSPCMIEHQETQK
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ RLSEVWQKVSQQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHRYPPSSSEE
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

```

SEQ CEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELDLHGL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ REETSAHIERKDKDITILQCLRLQELQLEFTETQKLTLLKDKFLQEKDEMLQLEKKLTQV
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCC.....

```

```

SEQ QNSLLKKEKELEKQQCMALEMTVKEAKQDKSKEAECKALQAEVQKLKNSLEAKQOER
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ LAAQQAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKSSMAEK
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCC.....

```

```

SEQ EQTSNRKRVEELSLELSEALRKLENSDKERQLQKTVAEQDMKMDMLDRIKHQHREQGS
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ IKCKLEEDLQEA TKLEDKREQLKKSKEHEKLEMEGELEALRQEFKKDKTLKENSRLKEE
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCC.....

```

```

SEQ ENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKH
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ YLQTTITKEAYDALSRSAACQDDLQALEKLNHVTSETKSLQOSLTQTEKKAQLEEEI
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ IAYEERMKKLNLRLKRGFQHESELEVHAFDKKLEEMSCQVLQWQKHQNDKMLAAKE
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCC.....

```

```

SEQ EGLREFQEEAALKENLLEDDKEPCCLPQWVSPKDTCLRYGNDQIMTNLEQWAKQQKVA
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```



```
COILS .....  
SEQ      NEKLGNGLREQVNYIAKLSEGDHLHSVMVHGLQQENKKLKEIEEKMKAEATRLCTKAL  
SEG      .....XXXXXXXXXXXXXXXXXXXXX  
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh  
COILS    ..CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
```

```
SEQ      GPSRTESIQREKVCVTGLGKWLPQDMGRMDLTTRYIGMHPHGSSYC  
SEG      .....  
PRD      cchhhhhhhhhhhhhhhhhcccccccccchhhhhheeecccccccccc  
COILS    ..
```

Prosite for DKFZphtes3 1g13.1

PS000001	52->56	ASN GLYCOSYLATION	PDOC000001
PS000001	68->688	ASN GLYCOSYLATION	PDOC000001
PS000004	240->244	CAMP PHOSPHO SITE	PDOC000004
PS000004	415->419	CAMP PHOSPHO SITE	PDOC000004
PS000005	74->777	PKC PHOSPHO SITE	PDOC000005
PS000005	110->113	PKC PHOSPHO SITE	PDOC000005
PS000005	238->241	PKC PHOSPHO SITE	PDOC000005
PS000005	290->293	PKC PHOSPHO SITE	PDOC000005
PS000005	392->395	PKC PHOSPHO SITE	PDOC000005
PS000005	396->399	PKC PHOSPHO SITE	PDOC000005
PS000005	444->447	PKC PHOSPHO SITE	PDOC000005
PS000005	503->506	PKC PHOSPHO SITE	PDOC000005
PS000005	544->547	PKC PHOSPHO SITE	PDOC000005
PS000005	566->569	PKC PHOSPHO SITE	PDOC000005
PS000005	600->603	PKC PHOSPHO SITE	PDOC000005
PS000005	650->653	PKC PHOSPHO SITE	PDOC000005
PS000005	655->658	PKC PHOSPHO SITE	PDOC000005
PS000005	735->738	PKC PHOSPHO SITE	PDOC000005
PS000005	876->879	PKC PHOSPHO SITE	PDOC000005
PS000005	968->971	PKC PHOSPHO SITE	PDOC000005
PS000006	39->43	CK2 PHOSPHO SITE	PDOC000006
PS000006	53->57	CK2 PHOSPHO SITE	PDOC000006
PS000006	68->72	CK2 PHOSPHO SITE	PDOC000006
PS000006	116->120	CK2 PHOSPHO SITE	PDOC000006
PS000006	190->194	CK2 PHOSPHO SITE	PDOC000006
PS000006	250->254	CK2 PHOSPHO SITE	PDOC000006
PS000006	296->300	CK2 PHOSPHO SITE	PDOC000006
PS000006	439->443	CK2 PHOSPHO SITE	PDOC000006
PS000006	444->448	CK2 PHOSPHO SITE	PDOC000006
PS000006	471->475	CK2 PHOSPHO SITE	PDOC000006
PS000006	520->524	CK2 PHOSPHO SITE	PDOC000006
PS000006	536->540	CK2 PHOSPHO SITE	PDOC000006
PS000006	566->570	CK2 PHOSPHO SITE	PDOC000006
PS000006	576->580	CK2 PHOSPHO SITE	PDOC000006
PS000006	650->654	CK2 PHOSPHO SITE	PDOC000006
PS000006	674->678	CK2 PHOSPHO SITE	PDOC000006
PS000006	804->808	CK2 PHOSPHO SITE	PDOC000006
PS000006	888->892	CK2 PHOSPHO SITE	PDOC000006
PS000006	963->967	CK2 PHOSPHO SITE	PDOC000006
PS000006	968->972	CK2 PHOSPHO SITE	PDOC000006
PS000007	135->143	TYR PHOSPHO SITE	PDOC000007
PS000008	202->213	MYR PHOSPHO SITE	PDOC000008
PS000008	599->605	MYRISTYL	PDOC000008
PS000029	83->105	LEUCINE ZIPPER	PDOC000029
PS000029	90->112	LEUCINE ZIPPER	PDOC000029
PS000029	97->119	LEUCINE ZIPPER	PDOC000029
PS000029	104->126	LEUCINE ZIPPER	PDOC000029
PS000029	403->425	LEUCINE ZIPPER	PDOC000029
PS000029	410->432	LEUCINE ZIPPER	PDOC000029
PS000029	918->940	LEUCINE ZIPPER	PDOC000029

(No Pfam data available for DKFZphtes3_ig13.1)

DKFZphtes3_1k11

group: cell structure and motility

DKF2phtes3.lkl1 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```

1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGCGC GCGGCGCGGG GGATCGTCTC
101 CGGCACTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGCG
151 GAGCAGCAGC GGTCCATGTA ACGTCAACCTT CTTCACAAAG GCCTCCACCC
201 CGGACTGTGT GCTGGCCAC CTTCAACACG CTTCGCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTGGGC GGGCGACCGT GCCTTCCCTT GTACCCGTGC
301 CGTGCTGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTT AGCCATGGCC
351 TTCGGGAGAG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCTCAC GCATCGCCAT
451 CAACGAGGAG AACGCTGAGT CACTGTGGA GCGAGCGGAC ATGCTGCAAT
501 TCCACCATGT GCGGGATGCT GCGGCGGAGT TCTTGAGAA GAACCTTTTC
551 CCTTCAACT GCTTGGCAT GATGCTCTC TCGGAGGCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCTT GCGCATGTG CTTGGTGCAC TTTGAGACGG
651 TGAGGCAGAG CGAGGACTTC AACAGCTGT CCAAGGACAC ACTGTGGAC
701 CTCATCTCGA GTGATGAGT GGAGACCGAG GACGAGCGGG TGGTCTTGA
751 GGCCATCCTC CAGTGGGTGA AGCAGCACT GGAGCCAGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGAGC GTGCGTCTGG CTTGTGTC GTCCGACTGC
851 CTGCAAGAGG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACGAGG ATCTGCGAGA
951 ATGATGGCGT GGTCAACAGC CCCTGTGCCC GGCCACGCAA GCGGGGCCAC
1001 ACGCTACTCA TCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACAC AAGGCCAAGG AGATCATCCC CAAGCCGAGC CTGCCAGGCC
1101 CCGGGAAGGA GTTCAGGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GTCGAGAA CCGGGTCTCC AAGGATGCTT GGGTGTACGA
1201 CACCGTACAT GAGGAATGGT CCAAGCGCGC GCCATGCTGT ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTGCCCGGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGTTGGGCC
1401 CCTTGGCGGA TGGCGTCAAG AATGCCCGAG TGGTGAAGTC CAAGCTGAAG
1451 CTCTTTGTTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAAGTGTAT GACCCCTCGG AGAACAGGTG GACGATCAAG GCGGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCGCTGCGG TCCTGGGCGC CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTCACAGCC GCTCGGCTT ACCGCTTTGA
1651 CTGTGAGACC AACCACTGGA CCGGATGAG GAGCATGACT GCGAAGGCGA
1701 TGCTCTGCCA TCCCTGCTGT TCCGCAACA AGCTCTATGT GGTGCGGGGC
1751 TACTTTGGGA CCGAGAGGTG TAAGACTCTG GACTGTATG ACCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTGCC CTACTCACTT ATCCCCACGG
1851 CCTTTGTGAG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCAAGCCAGA CCGCGCCCTT CAGTGTCA CA GGTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGGTGGAAGG
2001 GGCCTTGCCA GCTCTGGGSA GCAGCAGCCT TGGGCTGTTT TGAGCTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGCATCCG TGCCCTGGG GGCCTTTCA
2101 GCTTTGAGT GGTTTGTGGG AAGACATACC TCCAGAGGG GCATGGACTG
2151 CCACCGAGAC TGACCTGGC GTGCGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACCTG TTTGGCAGGT CTTGGACTGG GCGCGGCGAG GCAGGGGCGA
2251 GGAGGCGGCC CGGGTGGGCT TTTGGGCTGC GGCATGCCA CACATCCTTT
2301 CCCTCCTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGTG
2351 CTGGGCTGG GAAACTAGGT TCCAGGGGTT TGAGACAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAACTGCA GGGGGGCTC AGTGACATCC
2451 ATGAGGCTTT ATTAGCAAG GACACCCAGA CCTCAAGGT TTTGGGGCCC
2501 CTTCACAAA GCTGTAAGTC CAGACCCACC TACTCAGGGC CTGTCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGCCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCCTCCCT
2651 CCCCTCTCA GAGCCACCC TGAGAGGAGC CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GTTCTCCCC AGGAGACCAC GGTGCCCTCC

```

```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCTT CACCATCCCTT TTGGGTTTGG GCCCCATTG GTCCTCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCAATTGC
2951 TGTCACTCTC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTGGCGC CCTGTCACTG ACCATCCTAA TGCTCTCGC
3051 TCACTCCCAA GCCACCATTT GAGAGGGAGG GGTGTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC GCGAGGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCCTC TTGGGGGTGG GACCTTTGGG CAGGTTTTCG CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCCTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTGTCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

98350113:

Cloning of human ENC-1 and evaluation of its expression and regulation in nervous system tumors.

97252647:

ENC-1: a novel mammalian kelch-related gene specifically expressed in the nervous system encodes an actin-binding protein.

98234394:

NRP/B, a novel nuclear matrix protein, associates with p110(RB) and is involved in neuronal differentiation.

Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589
Category: strong similarity to known protein
Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLTNR KCMFTDVTIL
51 WAGDRAFFCH RAVLAASSRY FEAMFSGHLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAVSSRI AINEENAESL LEAGDMLQFH DVRDAAAEFL EKNLFPSNCL
151 GMMLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK OTLLDLISSD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEAV
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGQTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVYTGGRGS
351 ENGVSKDVMV YDTVHEWSK AAPMLIAREG HGSAELENCL YVVGHTSLA
401 GVFPASPSVS LKQVKEYDPG ANKWMHVAFL RDGVSNAVAV SAKLKLFEFG
451 GTSIHRDMVS KVQCYDPSEN RWTIKAECPQ PWRVYAAAVL GSQIFIMGCD
501 TEPTAASAYR FDCETNQWTR IGDHTAKRMS CHALASQNLK YVVGGYFGTQ
551 RCKTLDICYD TSQTNWCITT VPYSLIPTAF VSTWKHLPA

```

BLASTP hits

Entry MMU65079_1 from database TREMBL:

gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds.

Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611_1 from database TREMBLNEW:

gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens
nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.

Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314_1 from database TREMBL:

gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,
complete cds.

Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN). >TREMBL:DMRCPA_1 product: "ring canal protein"; Drosophila melanogaster ring canal protein and ORF2 mRNA, complete cds.
Score = 672, P = 3.9e-66, identities = 168/536, positives = 257/536

Alert BLASTP hits for DKFZphtes3_1k11, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_1k11, frame 2

Report for DKFZphtes3_1k11.2

(LENGTH) 589
(MW) 65923.45
(PI) 6.10
(HOMOL) TREMBL:MMU65079_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds. 0.0
(FUNCAT) 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
2e-09
(BLOCKS) BL01016D Glycoprotease family proteins
(PIRKW) zinc finger 1e-08
(PIRKW) DNA binding 1e-08
(PIRKW) transcription factor 1e-08
(SUPFAM) POZ domain homology 3e-68
(SUPFAM) vaccinia virus 59K HindIII-C protein 1e-15
(SUPFAM) A55R protein 5e-29
(SUPFAM) hypothetical protein YHR158c 4e-08
(SUPFAM) A55R protein middle region homology 5e-29
(SUPFAM) myxoma virus M9-R protein 1e-14
(SUPFAM) A55R protein carboxyl-terminal homology 5e-29
(KW) Alpha_Beta

SEQ MSVSVHETKRSSTGSMNVTLFHKASHPCVLAHLNLTLRKHCMTDVTWAGDRAFPCH
PRD ccc

SEQ RAVLAASSRYFEAMFSGHRESRDDTVNFQDNLHPEVLELLDFAYSSRIATNEENAESL
PRD hcc

SEQ LEAGDMLQFHQVDAADAAEFLEKNLFPSNCLGMMLLSDAHQCRRLYEFSSWRMCLVHFETVR
PRD nhhh

SEQ QSEDFNSLSKDTLLDLISSDELETEDERVVFETAILQWVKHOLEPRKVHLPPELLRSVRLAL
PRD hhh

SEQ LPSCQLQEAVSSEALLMADERTKLIMDEALRCKTRILQNDGVVTPCARPRKAGHTLLIL
PRD ccchhh

SEQ GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVMV
PRD ccc

SEQ YDTVHEEWSKAAPMLIARFGHGSAELENCLYVVGHTSLAGVFPASPVSLEKQVEKYDPG
PRD ccc

SEQ ANKMMMVAPLRDGVSNAAVVS AKLKL FVGGTSIHRDMVSKVQCYDPSENRTWIKAECPQ
PRD ccc

SEQ PWRYTAAAVLGSQIFIMGDTFTPTAASAYRFDCTNOWTRIGDMTAKRMSCHALASGNKL
PRD ccc

SEQ YVVGGYFGTQRCKTLDCYDPTSDTWNCITTVPSLIPTAFVSTWKHLPA
PRD ecc

(No Prosite data available for DKFZphtes3_1k11.2)

(No Pfam data available for DKFZphtes3_1k11.2)

DKFZphtes3_in3

group: signal transduction

DKFZphtes3_in3 encodes a novel 1196 amino acid protein with similarity to S. pombe Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```
1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAAACTTG TCAGGTCTGA AGAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACACAAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCATATT AAAGAACTA
251 CAAGTGATGA TGTAAAGTCT GCTAACACTA ACAACCTGAA GAAGAGCAGC
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAAAATAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGATT TGATGCAAGC ATATCAGTGC
601 CATGTAACTG AAGAAATGGC AAAGGAGATT AAGAGGAAA TAAGAAAGAA
651 ACTGAAAGAA CAGTTGACTT ACTTTCCCTC AGTACTTTA TTCCATCATG
701 ACAAACTAAG CAGTGAAAAA AGGAAAGAGA AAAAGGAAGT TCCAGTCTTC
751 TCTAAAGCTG AAACAGGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
801 TGAAACAAGG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGCTCA ATGGAACAAA GCACAGAAGA CAGCATGCAA
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAGAAGA CTAAAGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTCAT GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTACCGC AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCAATGTGGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTCTCTTCA CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTGTGCT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAAAT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAAATG CAACATCAAA CTCAAAACTT
1501 CGCTTGAGCG TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAAGTG
1551 TGTGAGGCA TTTGAATGCT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACGTGA AGAGGACTGA AAGTTCCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAAG GTAAACCAAG
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGG AAGCACTCCC TGGCGAGCT
1801 TGCCGATATCC CAAACAAACA CCTCTTCTCA CTAATGTCAG GAGAACGAGG
1851 ATGTTTTTGT CTGATTCTCT CCCACATGG AAGAATATTA GCAGCAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTAATT TATATGAAAT TCCTCTGGGA
1951 CTTTTCATGA GAGATTTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGCTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACCTT CAGAGTTTTA
2101 CCTCATCCTT CTTTTGTTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTTGATACT GAAGGTCATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAAATTAAG
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTGGAGA TTCAATCCAA
2451 TGGAAAACGT TTGTTAATCC ATACCAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTG ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTCGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGAA CCCAGAAACA GGAGAACAAG
```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTGG AGACATTCTT
2701 TATCATCCAT TTGAAAATAT GGTTCATTC TGTCATTGG GCAAAATGA
2751 GCCAATTCCT CTGTATATTT AGGATTTCCA TGTTGCCAG CAGGAGGCTG
2801 AATGTTCAA ACCTTACAAT GGACATTTT CATTACCTGG AATACACCAA
2851 AGTCAAGATG CCCTATGTAT CTGTCCAAAT CTACCCCATC AAGGCTCTTT
2901 TCAGATTGAT GAATTTGTCC ACACCTGAAAG TTCTTCAACG AAGATGCAGC
2951 TAGTAAACA GAGGCTTGA AACTGTACAG AGGTGATACG TTCTGTGCT
3001 GCAAAAGTCA ACAAATCT CTCTTTACT TCACCACAG CAGTTTCTCT
3051 ACAACAGTCT AAGTTAAAG AGTCAACAT GCTGACCGCT CAAGAGATTCT
3101 TACATCAGTT TGGTTTCACT CAGACCGGA TTATCAGCAT AGAAGAAAG
3151 CCTTGTAACT ATCAGGTAGA TACAGCACCA ACGGTAGTGG CTCTTTATGA
3201 CTACACAGCG AATCGATCAG ATGAACATAC CATCCATCGG GGAGACATTA
3251 TCCGAGTGT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAAACACT
3351 GTATCAAGAA CTGCTCCTG AGATAAAGGA GCGATCCCTC CCTTTAAGCC
3401 CTGAGGAAAA AACTAAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA
3451 ATCAATAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC
3501 ACATTCTGAA ATGAGAAAG AACAGAGCCA TGAGGACCAA GGACACATAA
3551 TGGATACAGG GATGAGGAAG AACAGCAAG CAGGCAGAAA AGTCACTCTA
3601 ATAGAGTAAA GAATTGAAGA AAAGTTAAGA GCTGCCGAAA TGCACAGAGG
3651 TGAATAAGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCTG
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTTGTCTTT ATGAATGACT
3751 CTAGAAAAAT CAGAAATCAAG TTGTGGGTGG AAAATCAAC GTGGCTTTTG
3801 AGTTCAGTGG TTATAAACC TGTGACTAT TGTGGTCAA AGTATTGTTA
3851 CTATATTGT TAGTAATGC ATCATAATTA CATTACCAGT GTTGGAAAC
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAATGTGA ATAAAGGTG
3951 TTTGCTTAT TAGGATGCT GTTAAGTAAT CATTAAATAT TATTATATTG
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT
4051 TGAATTCAC TTTATTAAA AGATAAGCAG CTGACTGGGC ACGGTGCTCT
4101 ATGCTGTAA TCCTAGCACC TTGGGAGGCT GAGGCAGGTG GATCACCTAA
4151 GGTGAGGAGT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT
4201 CTACTAAAAA TACAATAATC AGCCGGGCTC CTGGCAGGC ACCTGTAATC
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTGCAG
4301 TGAGCCAAGA TCACGCCATT GCACCTCCAG CTGGGGGACA GAGCAAGACT
4351 CTATCTCAA AAAACAAAAA AGATAAGCAG CTTAGAAATA TGGCGCATTC
4401 AAAACAGTCT CAGTAAACAA GACATTAAAA GAAAAAATT TACTTTCTAA
4451 TTAATAATTT GTGTTCTTAA AGATCAAAAT ATATAGGTAA CTTATAGAC
4501 CTAATAATAA AGTGATTTT GGCTGGAGCT GCAACAATG TCCCAATGTC
4551 TTTACTTTTT AAAAAGGCT TTTATATTT AAGCACATAC CTATTTTGT
4601 GACTTACATT GTTTAATATT TATTTAATC TTAATATTT TACATTATTA
4651 TATTCATTA TTTATTTTT CTAGTTTCCA GAATAATGT GCATTTATTA
4701 TAGACTATAT GTTTTGAAT TGTATATTT AATGGATAT TCATTTTTTG
4751 TTTCTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATAAAA
4801 TATTTAACC CAAGAAAGTG AAAACTAATA TAAATTTAGA AAGACCTATC
4851 CAAATAGAC AGTCAATTC ATTAATAATA AAGTGAGAA AACAATGTT
4901 GGGCATTGAG GTGTAAATTT TGCCAGATG TATACCCAGT GTGAAATATC
4951 TTCTAATAAA AATATATTT GCTCTTATCC CTGCACATGT AGAGGCATAA
5001 AAATTGGTAA ACATGTCCTG CTGTGTAGAA CTTTAAAAA AAGGCATTTT
5051 TGAAGTGTG GAGTGGCCT GATAACTGGT GAAGCCTACA GCCATCCGCC
5101 CAAAAGCTG TTCTGATGCG ACTGAGTTTT CATTGTTCTG GATGTATAAG
5151 TCTGTGTGCT AGGTACAGCT GGGCCAGGCC AGCTTGAGTC ACTCTGTGAC
5201 AAGCTTGTTT TTTCTGCT TGTGAATGCA CTTGATAATT TAAAAATAAA
5251 AATATCTGTT TCTCTGAAA AAAAAA

BLAST Results

Entry HS3281 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 32B1
Score = 4445, P = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:
Human exon-trapped sequence from 6q24.
Score = 965, P = 4.0e-35, identities = 193/193

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196
Category: similarity to known protein

```

1 MPTAESEAKV KTKVRFKELL KTHSDLMREK KKLKKLVRS EENISPDITR
51 SNLHYMKETT SDDPDTIRSN LPHIKETTSO DVSAANTNNL KKSTRVTNKK
101 LRNTQLATEN PNGDASVEED KQKPNKKVI KTVPLTTOD LKPETPENKV
151 DSTHQKTHTK PQPGVDHQS EKANEGREET DLEDEELMQ AYQCHVTEEM
201 AKEIKRKIRK KLKEQLTYFP SOTLFHDDKL SSEKRRKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHDD EISSMEQSTE DSMQDDTKPK
301 PKTKKKKTKA VADNNEDVDG DGVEITSRD SPVYFKCLLD DDLVLGVYIH
351 RTDRLKSDFM ISHPMKIHW VDEHTGQYVK KDDSGRPVSS YYEKENVDYI
401 LPIMTPQYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDES PKVILFFEIL
451 DFLSVDEIKN NSEVQNEQCG FRKIWAFLK LLGANGNANI NSKLRLQLYY
501 PPTKPRSPLS VVEAFWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCERHHESS SVOTEPGLEE SKEVIKWKRL PGQACRIPNK
601 HFLFSNAGER GCFCLDFSHN GRILAAACAS RDGYPIIYE IPSGRFMREL
651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
701 YTAKEHFAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CFDTGHHMY SGCCTGVIVV WNTYVKINDL EHSVHHWTIN KEIKETEFKG
801 IPISYLEIHP NGKRLLIHTK DSTLRIMDLR ILVAKFVGA ANYREKIHST
851 LTPCGTFLFA GSEGGIVYVW NPETGEQVAM YSCLFFKSP1 RDISYHPFN
901 MVAFCAGQW EPILLYTYDF KVAQQAEMF KRYNGTFPLP GIHQSQDALC
951 TCRKLPKQGS FOIDEFVHTE SSSTKMLVK QLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQSKLKQSN MLTAQELHQ FGFTGTGIS IERKPCNHQV
1051 DTAPTVALY DYTANRSEDL TIHRGDIRV FFKDNEDWY GSIGKQGEY
1101 FPAHVASET LYQELPPEIK ERSPLSPEE KTKIEKSPAP OKQSINKNKS
1151 QDFRLGSESM THSEMRKES HEDQGHIMDT RMKKNKQAGR KVTLIE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_in3, frame 1

TREMBL:U92792_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CERO3E1_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:YZLL_CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds.
Length = 321

HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18
Identities = 59/225 (26%), Positives = 111/225 (49%)

```

Query: 647 MREL CGHLNIIYDL SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFH 706
      + E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +F+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD--LLSASMDKTVRLW--QVGRDSCLKVFSTHYVTCVQFN 119

Query: 707 PAVREL VVTGCYDS MIRIWKVEMREDSAILVRQF DVHKSFINSLCFDTGHHMYS GDCGTG 766
      +TGC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITG CIDGLVRIMDVK-----CLVVDWMSKEIVTAVCYRPGDKGAVAGTITG 174

Query: 767 VIVVWNTYVKINDLEHSVHHWTINKKEIKETEFKIPISYLEIHPNGKRLLIHTKDSTLRI 826
      ++ +LE V ++N K ++ Y P K+L++ + D+ +RI

```

Sbjct: 175 NCRYYDASENRLEESQV---SLNGRKKSLHKRIVGFQYCPSPD--KKLMVTSGDAQVRI 229

Query: 827 MDRLILVARKEVGAANYREKIHSTLTPCGTFLFAGSEGDGIVYVMN 871

Sbjct: 230 LGGAHVSN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIMYMN 272

Pedant information for DKFZphtes3_ln3, frame 1

Report for DKFZphtes3_ln3.1

[LENGTH] 1196
[MW] 137114.70
[pI] 6.79
[HOMOL] SWISSPROT:YKY4_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN C14B1.4 IN CHROMOSOME III. 8e-21
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-08
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 9e-08
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 9e-08
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06
[FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 4e-06
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 4e-05
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05
[BLOCKS] BL00024H
[SCOP] dltbqd_2.46.3.1.1 betal-subunit of the signal-transducing 3e-91
[SCOP] dlqfc_2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14
[SCOP] dlrmk_2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15
[SCOP] dladd5l_2.21.2.1.7 (1-63) Hemopoietic cell kinase Hck [human (Hom 3e-15
[SCOP] dlckal_2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13
[SCOP] dlqwea_2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15
[SCOP] dlshg_2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13
[SCOP] dlprmc_2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15
[SCOP] dlhsq_2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13
[SCOP] dlboa_2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13
[SCOP] dlfnh_2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15
[SCOP] dlsema_2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13
[SCOP] dlgbqa_2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16
[SCOP] dlckaa_2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15
[EC] 3.1.4.3 Phospholipase C 2e-07
[EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07
[EC] 3.6.1.32 Myosin ATPase 7e-07
[EC] 2.7.1.112 Protein-tyrosine kinase 8e-06
[PIRKW] nucleus 2e-08
[PIRKW] phosphotransferase 8e-06
[PIRKW] plasma 4e-07
[PIRKW] duplication 4e-07
[PIRKW] phosphoric diester hydrolase 2e-07
[PIRKW] tandem repeat 7e-07
[PIRKW] hormone 4e-07
[PIRKW] transmembrane protein 2e-06
[PIRKW] stomach 4e-07
[PIRKW] actin binding 7e-07
[PIRKW] ATP 7e-07
[PIRKW] phosphoprotein 7e-07
[PIRKW] signal transduction 7e-09
[PIRKW] heterotrimer 7e-09
[PIRKW] P-loop 7e-07
[PIRKW] hydrolase 7e-07
[PIRKW] transcription regulation 5e-06
[PIRKW] GTP binding 7e-09

[SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07
 [SUPFAM] SH3 homology 2e-07
 [SUPFAM] SH2 homology 2e-07
 [SUPFAM] protozoan myosin heavy chain IB 7e-07
 [SUPFAM] myosin motor domain homology 7e-07
 [SUPFAM] pleckstrin repeat homology 2e-07
 [SUPFAM] protein-tyrosine kinase src 8e-06
 [SUPFAM] WD repeat homology 3e-12
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07
 [SUPFAM] protein kinase homology 8e-06
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-09
 [SUPFAM] yeast coatomer complex alpha chain 4e-07
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 19
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] Src homology domain 3
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW COMPLEXITY 5.77 %
 [KW] COILED_COIL 2.42 %

SEQ MPTAESEAKVTKVRFKLLKTHSDLMREKKKKKKLVRSEENISPDITIRSNLHYMKETT
 SEGXXXXXXXXX.....
 COILSCCCCCCCCCCCCCCCCCCCCCCCC.....
 lgotB

SEQ SDDPDTIRSNLPHIKETTSDDVSAANTNNLKKSTRVTNKNLRLNTQLATENPNGDASVEED
 SEG
 COILS
 lgotB

SEQ KQKPNKKVKTVPOLTTQDLKPETPENKVDSTHQKTHTKPOPGVDHQKSEKANEGREET
 SEGXXX
 COILS
 lgotB

SEQ DLEDEELMQAYQCHVTEENAKEIKRKIRKKLKEQLTYFPSDTLFHDDKLSSEKRRKKKKE
 SEGXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXX
 COILS
 lgotB

SEQ VPFVSKAETSTLTISGDTVEGEQKKFSSVRSVSSDSHQDDEISSMEQSTEDSMQDDTKPK
 SEGXXXXXXXXX.....XXX
 COILS
 lgotB

SEQ PKKTKKKTRAVADNEDVDGQVHEITSRDSVPYPKCLDDDLVLGVYIHRDLRLKSDFM
 SEGXXXXXXXXX.....
 COILS
 lgotB

SEQ ISHPMVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTQPYDFKQLKSRLPEW
 SEG
 COILS
 lgotB

SEQ EEQIVFNENFPYLLRGSDSPKVILFFEILDLSVDEIKNNSEVQNQECGFRKIAMAFK
 SEG
 COILS
 lgotB

SEQ LLGANGNANINSKLRLQLYYPPTKPRSPLSVVEAFEMWSKCPRNHYPTLYTVTRGLKVP
 SEG
 COILS
 lgotB

SEQ DCIKPSYRSMALQEEKGKPVHCERHHESSSVOTEPGLEESKEVIKWKRLPGQACRIPNK
 SEG
 COILS
 lgotB

```

SEQ      HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPIILYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECCCCCEEE

SEQ      SWSKDDHYILTSSDGTARIWKNEINNTNTRVLPHPSFVYAKFHPAVRELVTTCYDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEETTTEEEETT--TTCEEEETTTCCEEEETT--TCCEEEETT

SEQ      MIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEEETTTTBTTEEEEEECCCE-EEEEETTEEEETTTEEEEEE.....

SEQ      EHSVHHWTINKETEFKGIPIISYLEIHPNGKRLLINTKDSLRLMDLRLVARKEVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLFAGSEGDIVVWNPETGEQVAMYSOLFPKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAFQGNPILLYIDFVHAQQAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FQIDFVHTESSTKMQLVKORLETVEVIRSCAAVKNKLSFTSPPAVSSQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQEILHQFGFTQTGTIISIERKPCNHQVDTAPTVALYDYTANRSDTLTHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPPLSPEEKTKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      QKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTMRKNKQAGRKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

Prosites for DKF2phtes3_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PD0C00001
PS00001	686->690	ASN_GLYCOSYLATION	PD0C00001
PS00001	934->938	ASN_GLYCOSYLATION	PD0C00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PD0C00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PD0C00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PD0C00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	48->51	PKC_PHOSPHO_SITE	PD0C00005
PS00005	66->69	PKC_PHOSPHO_SITE	PD0C00005
PS00005	93->96	PKC_PHOSPHO_SITE	PD0C00005
PS00005	170->173	PKC_PHOSPHO_SITE	PD0C00005
PS00005	232->235	PKC_PHOSPHO_SITE	PD0C00005
PS00005	268->271	PKC_PHOSPHO_SITE	PD0C00005
PS00005	304->307	PKC_PHOSPHO_SITE	PD0C00005
PS00005	327->330	PKC_PHOSPHO_SITE	PD0C00005
PS00005	352->355	PKC_PHOSPHO_SITE	PD0C00005
PS00005	384->387	PKC_PHOSPHO_SITE	PD0C00005
PS00005	440->443	PKC_PHOSPHO_SITE	PD0C00005
PS00005	533->536	PKC_PHOSPHO_SITE	PD0C00005
PS00005	546->549	PKC_PHOSPHO_SITE	PD0C00005
PS00005	643->646	PKC_PHOSPHO_SITE	PD0C00005
PS00005	677->680	PKC_PHOSPHO_SITE	PD0C00005
PS00005	690->693	PKC_PHOSPHO_SITE	PD0C00005
PS00005	702->705	PKC_PHOSPHO_SITE	PD0C00005

PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

Pfam for DKFZphtes3_1n3.1

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*		
Query	650	LCGHLNIIYDLSWSKDDHY-ILTSSSDGTARIWK	682
HMM_NAME	Src homology domain 3		
HMM	*pyVIALYDYqAqdpDELSFkEGDIIiIIEdsDD.WMrqRnnnTNGQEGW		
	P+V+ALYDY+A+++DEL++ +GDII + +++ WM+G GQEG+		
Query	1054	PTVVALYDYTANRSDELTIHRGDIIRVFFKDNEWYGSIGK--GQEGY	1100
HMM	IPSNYVEPI*		
	+P+N V+ +		
Query	1101	FPANHVAE	1109

DKFZphtes3_20c21

group: testes derived

DKFZphtes3_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
1 GGTAGGCGGG GCGGGCGGTG ACCTAAGGCC TCTCTGCCGC GCGGCGAGGT
51 ACGGGGCGAGA AGTCGCAGGT ACCCAGCTGC TGCCACGTTT TCTGGTCCAG
101 AGTCCCGAAC CCCGAGCACT GGGATGCGCT GCTACTCCGA GCCAAGCCAC
151 TGATGTTTGA ACTGGAAACT TCAAAACGTT TAATAAGACT CTTCAGGATG
201 GGTTCGAACT AGACAAGCTA GAAATTTCCT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTCTTGCTCT TCCTGGAGAG GAGCTTGAAG AGGTGCTCT
301 GCAGACAGCC ACAGTGATAC TCAGGAACCC AGAGGAATGG ATTTGACTTT
351 TCTGTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTGTGA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACTTTCC
451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCAAGT CCAGATTAC TGGATTGGGG GATAGAGGAA CGGTACACGC
551 TAGGAGAGAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAAGTCA
651 GCCTCGTGGT GGAATTATTT TTTCTTTTAT GATGGTTCCA AGGTAAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTGG TTACTTTTAT CCTTCCACGA
751 CCCTGTAGTA CCAACAGGAG TTGCTTTGTG GACAGATTGC TGGAGTTGTC
801 CGCTGTGTTT CTGACATTTC TGACTCTCTCT CTTACTCTTG TTGCTGTGAG
851 AAAACTGAAG TTGGCATAAA AAGTTGATGG AGATTACCTG TGGGTGCTGG
901 GCTGTGCTGT CGAGTCTCCT GATGTCAGCT CCAAGCGGTT TCTGGATCAG
951 CTAGTTGGAT TCTTTATTTT TTACAATGGA CTGTGTTCCC TAGCTATGA
1001 GAAGTGTCTC CAGGAAGAAG TGAGCACGGA GTGGGACACC TTCTCGAGC
1051 AAATTCTGAA AAACACCACT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAATAAAGT GGAGCCCCTG TTGTTGCTGA AGGCAGCCCG
1151 CATTCTGCAG ACCTGCCAGC GCTCGCCTCA CATTCTCGCT GGCTGCATCC
1201 TCTATAAAGG ACTGATTGTC AGCACCCCAAC TCCGCCCCTC CCTCACCGCC
1251 AAGGTCTCTG TTCAACGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTCAGCC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAAG
1501 AAAACGCCAC TGGCCATGTT GAATCCATGG CTTGGACCAC CCCAGATCCC
1551 ACATCCCTCT ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACGT CACAACTCTG
1651 CCAGGGGTGA GGTCTCTGGC CTCAGCTCCT CCTCGGGGAA GGAAGTAGTC
1701 TTTCTCAAG AAGAACTCGA CTTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGAA ATGGCCTCAG GTCAATTTGC TTCTCTACAT GTGCCTGTTT
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCCAGCAGC
1851 CTGGAACCCA CGCTCTCTGA GGACACAGCC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTGGAAG
1951 ACCATCTCTG CCATAGCAGC CAAGCCCCCA TTCCAGAGC AGACCTCTCT
2001 CCCAGAAGGA CCGCAGGCC CTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGAAAC AAGCTTCCCA CGGGGGAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGCCAC GCAGCCCCTG GTCTGGAATG CAGTTCAGGC
2151 TCAGCAAACT GTCAGGGTGC TGGCCCCCTC GCAGATGGAA TCAGTCCAG
2201 GCTGACACCA GCAGAGTCTT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACTGCGT CAAAGGGCTG ATGCTGTCCC TGCTGGCTGA GGAGCCGCTG
2301 CTGGGAGACA GCGCAGCCAT AGAGGAAGTG TACCACAGCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAAGA GAGCTGCCC AGGATGAGG
2401 CAGCCTCCAC GAGCAGCAGC TACAATCTCA CATATTACGA CGCATTCAG
2451 AGCTTGCTGA TGGCAAACTT CCGCAGGTG GCCACCCCGC ATGATCGCCG
2501 CTCTCTCCAG CGCTCAGCC TGATGCAATG CGAATTTGCC CAGCTGCCCG
2551 CGCTTATGA AATGACTCTC AGAATGCCT CCACGGCTGT GTAGCCTGT
2601 TCCACCCCA TCCAGGAGC ATATTTCCAG CAGCTGGCAG CTGCAGCAGC
2651 GAGCTCCGGC TTCCCAACC CTCAGGATGG CGCTTCAGC CTCTCCGGCA
2701 AAGCAAGCA GAAGCTGCTG AAGCAGGGG TGAATCTGCT CTGAAGTGA
2751 CCCAGGAGT GACTGGGAAG GAGAAAACA GCAAAGGAAG CTCTGCCTTT
2801 TATAATTGAA AAGGCCCTCT TATTTTATTT TTCTTGAAAA CATTCCCTTT
```

```
2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG CTTAGTTCTT CAGTCCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACGTGATT TAATCGGTAG TGTTGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTCATA TCTCCTGGCT
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCCAAGTT GGCATTTTTT
3101 CTAGAACCTG ATCGTCCACT AGCCCAAGT GTCTGTGTC AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGCAGT AGGTTTCTCC AGTGGGCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTGG AATTCTGTGG GAACCTACT CAAATCTAGG GGCCTCTTCT
3351 TTGGACCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCTGC GTGCTTTTG ACTACAAGT GGCCTGGCAG
3501 CTGGTTCTCT TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGCTTG
3551 GGGGTAAAG CACTGTGCTT TTCACTGGTG CTGCGTGA ACGGAGCGAC
3601 ACTCAGCTGT GTGTTCTCTGG GCTTGTCTGG TACTTAGAAC CTCAGTTCTA
3651 TTACGTTATA CTCAGACATT TTTTTCAGC TATGAGACAG ACTGCAGGAT
3701 GAAATATTTT GTCAAAATCT TAACGAATG TTTACTGAA GTACTTGAGA
3751 TTCCATTGGA GAGTTGTATT GTTAATAAT TCATGTCAGT GAACGTGAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS1048E9 from database EMBLNEW:
Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2
Contains pseudogene similar to ribosomal protein S3A and part of a gene
similar to C.elegans protein CE02118, ESTs, STS, GSS.
Score = 6540, P = 0.0e+00, identities = 1308/1308
-14 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708
Category: putative protein
Classification: no clue

```
1 MATSTSTEAK SASWNNYFFL YDGSKVKEEG DPTRAGICYF YPSQTLLDQO
51 ELLCQGIAGV VRCVSDISDS PPTLVRLRKL KFAIKVDGDY LHWLGCVEL
101 PDVSKRFLD QLVGFNFYN GPVSLAYENC SQEELSTEND TFIEQILKNT
151 SDLHKIFNSL WNLQOTKVEP LLLKAAARIL QTCQRSFIL AGCILYKGLI
201 VSTQLPPLSL AKVLLHRTAP OEORLPTGGD APQEHGAALP PNVQIIPVTV
251 TKEEAILLHE FPVEQMTRSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWITPD PTPDEACPD GRKENGCLSG HDLESIRPAG LHSNARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPEAQEV EMASGHFAFL HVPVPOGRAP
401 YCKASLSASS SLEPTPPEDT AISSILPPSA PEMLTQHQAG EQVEDHPGHS
451 SQAPIPRADP LPRRTTRPRL LPRLDPGQRG NKLPTGEQGL DEDVDGVCE
501 HAAPGLECSS GSANCOGAGP SADGISSRLT PAESCMGLVR MNLYTHCVKG
551 LMLSLLAEPP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTTYDRI QSLLMANLPQ VATPHDRRL QAVSLMHSEF AQLPALYEMT
651 VRNASTAVYA CCNPIQETYP QQLAPAARSS GFNPQDQAF SLSGKAKQKL
701 LKHGVNLL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_20c21, frame 3

Report for DKFZphtes3_20c21.3

(LENGTH) 708
(MW) 76900.23
(pI) 5.30
(KW) Alpha_Beta
(KW) LOW_COMPLEXITY 6.36 %

SEQ MATSTSTEAKSASWNYFFLYDGSKVKEGDPTRAGICYFYPSQTLDDQQLLCGQIAGV
SEG .XXXXXXXXXXXXX.....
PRD ccc

SEQ VRCVSDISDSPPTLVRLRKLKFAIKVDGDLWVLGCAVELPDVSKRFLDQLVGFFNFYN
SEG
PRD eeeeecc

SEQ GPVSLAYENCSEELSTEWDTFIEQLKNTSDLHKIFNSLWNLDTKVEPLLLKAAAIL
SEG
PRD ccc

SEQ QTCQRSPHILAGCILYKGLIVSTQLPPSLTAKVLLHRTAPQERLPTGGDAPQEHGAALP
SEG
PRD hhhcc

SEQ PNQIIPVFVTKKEATSLHEFPVEQMTSLASAPAGLDGSAQHHPKGGSTSAKENATGH
SEG
PRD ccc

SEQ VESMAWTTDPDTPDEACPDGRKENGCLSGHDLISIRPAGLHNSARCEVLGLSSSLGKEL
SEG
PRD ccc

SEQ VFLQEELDLSEIHIEAQEVEMASGHFAFLHVPVFDGRAPYCKASLSASSLEPTPPEDT
SEG
PRD hhhhhhhcc

SEQ AISSLRPPSAPEMLTQHGAEQVEDHPGHSSQAPIPRADPLPRRTLRPLLPRLDPGQRG
SEG
PRD ccc

SEQ NKLPTEGQGLDEDVDGVCESHAAPGLECSSGSANCQAGPSADGSSRLTPAESCHGLVR
SEG
PRD ccc

SEQ MNLYTHCVKGLMLSLLAEPPLGDSAAIEEVYHSLASLNGLEVHLKETLPDEAASTSS
SEG
PRD ccc

SEQ TYNFTYYDRIQSLLMANLPQVATPHDRRFLQAVSLMHSEFAQLPALYEMTVRNASTAVYA
SEG
PRD ccc

SEQ CCNPIQETFFOQLAPAAARSSGFNPQDGAFLSGKAKQKLLKHGVNLL
SEG
PRD ecc

(No Prosite data available for DKF2phtes3_20c21.3)

(No Pfam data available for DKF2phtes3_20c21.3)

DKFZphtes3_20k2

group: signal transduction

DKFZphtes3_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CGATGACAGT
101 GTGATGGAGA GTCTCTGGCG TGCCATCTGG GATGCAAAAC GTCCCTGTGT
151 CCCCCACGTC CAGGCCGTAG ATGCTCCCCG CCGGTACGTC ACTTAGTCGT
201 CAGATCGCCC GTCTCGGTAT CACAGTGCCT CTGTTACAGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGGC GACCACTCC AAAAGGACAC CTGCCACAGC CCCCCTGGATG
351 GAGACCCCTAA CTCACGGCCA CCTCCAGCCA AGCCCCAGCT CTCACAGGCC
401 AAGAGCCGCA CCGGGCTCTT TGGGAAGGGT GACTCGGAGT AGGCTTTCCC
451 GGTGGATTGC CCTCACGAGG AAGGTGAGCT GGAATCCTGC CCGACCATCA
501 CAGTCAGCCC TGTATCACCC ATCCAGAGGC CAGGAGACGG CCCCACCGGT
551 GCCAGGCTGC TGTCCCAGGA CTCTGTCCGC GCCAGCACCG AGAAGACCCT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCGTGTCT CAGAATAACT
651 GCCAGATCTT GGAGAGCCTG CTGCTCTTCC TGCAGAGAGC CAAGAAGCAC
701 CTCACAGACA ACAGGTTCAA AGACCTGAGC ACAGGGAAGA CCTGTCTGCT
751 GAAAGCCATG CTCACCTGCG ATGACGGACA GAACACACCC ATCCCCCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
851 AGCTACACGG ACAGCTACTA CAAGGGCCAG ACAGCACTGC ACATCGCCAT
901 CGAGAGACGC AACATGGCCC TGGTGACCTT CCTGGTGAG AACGGAGCAG
951 ACGTCCAGGC TGGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTTG TCCCTGGCCG CGTGACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGCTGCA GAACTCCTGG CAGACGGCCG
1101 ACATCAGCGC CAGGGAATCG GTGGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACAGC AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCCTGG GGCCCAACT GCACCCGAGC CTGAAGCTGG
1251 AGGAAGCTAC CAACAAGAAG GGAATGACCC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGCTTT GGCCTATATT CTCCAGCGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTCGCTGTAC GACCTGTCTT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCATAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCTT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTTGCTCCCC TTTAAGATGG AAAAAATTGG AGACTATTTC CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTTT CCGAGGGATT
1751 CAGTATTTCC TGCAGAGGCG GCGCTGATG AAGACCTGT TTGTGGACAG
1801 CTACAGTAGA ATGCTTTTCT TTCTGACGTC ACTGTTCATG CTGGCCACCG
1851 TGGTCTGTGA CTTCAGCCAC CTCAGGAGAT ATGTGGCTTC CATGGTATTC
1901 TCCTGGCCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTGTCT TACATCGTCT TCTTGTTCGG GTTTTCCACA
2051 GCGGTGGTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTGC CACAGGTGGC GGGGGCCTGC CTGCAGGCCC CCGGATAGCT
2151 CCTACAACAG CCTGTACTCC ACCTGCCTGG AGCTGTTCAA GTTCACCATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCTG CTGCTGGCCT ATGTAATTCT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGAG
2351 ACBAAGACA TCTGGAGCTT GCAGAGAGCC ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCGCTCA GGCAAGCTGC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACCT GACCACTGGA AACACCAACG TGGGATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGCGGT CAAGCGCACG CTGAGCTTCT
2601 CCCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGGC
```

```
2701 CGAGGAAGTT TATCTGGGAC AGTTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTGAC
2801 GCAGACAGCA CTGTCAACAC TGGGCTTAG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACAGG TGCTCTGTGA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCGAGC ATGTTCCCAA ATCTGTGCTG GACAAGCTGT GGGAGCGGTT
2951 CTTGGAAGCA TGGGAGTGA TGTACATCA ACCGTCACGT TCCCAAGTGT
3001 AATCTCCTAA CAGACTTCA GCTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTCCT TTGATTTTA
3101 TTCTTTTTTT TGAGACAGAA TTTCACCTTT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCTGCGCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGTCTGGCTA ATTTTTTGTA TTTTTTAAT AGATATGGGG TTTCGCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATCTTTT TTTTTTTCT GTGAGACAGA GTTTCACCTT
3451 TGTGCCCAGC GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCTCTCCGG GTTCAAGCGA TCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGCT GAGCACTACC ACCGCCCGCT AATTTTTGTA TTTTAAATAG
3601 AGACGGGGTT TCACCATGTT GGCAGGCTG GTCTCGAATC CTGACCTCA
3651 GGTGATCTGC CGGCTTGGC CTCCCAAAGT GCTGGGATTA CAGGTCTGAG
3701 CGGCTGGCCT CGGCTTCTT TGATTTTATA TTATTAGGAG CAAAGTAAA
3751 TGAAGCCAG GAAACACCT TTGGGAACAA ACTCTCTCTT TGATGAAAA
3801 TGCAGAGGCC CTTCCTCTCT GTGCGTGTCT TGCTCTCTTT ACCTGCCCGG
3851 GTGGTTTGGG GGTGTTGGTG TTTCTCTCTT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTTGGCA GAATCAAGCT GTTGACAGCAG TGCTTCTTTC
3951 ATCCTTCTCT ACATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAAGATC TTCCATTCTT AGGAACCCCA GTCTGCTTTC TCCGCAATGG
4101 CACATGCTTC CACTCCATCC ATACTGGCAT CCTCAATTA ACAGATATGT
4151 ATACATATA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

99288727:
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:
A non-pungent triprenyl phenol of fungal origin, scutigerol, stimulates
rat dorsal root ganglion
neurons via interaction at vanilloid receptors.

Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839
Category: strong similarity to known protein
Classification: Cell signaling/communication

```
1 MKKWSSTDLG AAADPLQKDT CPDPLGDPN SRPPPAKPQL STAKSRTRLF
51 GKGDSSEAFV VDCPHEGEL DSCPTITVSP VITIQRPGDG PTGARLLSQD
101 SVAASSTKTL RLYDRSIE AVKQNNQDL ESLLLFLOKS KKHLDNEFK
151 DETGKTCLL KAMLNLDGGO NTIPLLEI ARQDLSKEL VNASVYDSYY
201 KGOTALHIAI EERNMALVTL LVENGADVQA AAHGDFFKT KGRPGFYFGE
251 LPLSLAACTN QLGIVKFLQ NSWQTADISA RDSVGNVVLH ALVEVADNTA
301 DNTKFVTSY NEILILGAKL HPTLKEELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECHLSRKF TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSETFNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 MAAYRFPVDG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFFLQS LFMLATVVLY FSHLKEYVAS MVFSLALGWT
551 NMLYYTRGFQ QMGIIYVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGNDSLPSF STSHRWGPA CRPPDSSVNS LYSTCLELFK FTIGMGDLEF
651 TENYDFKAVF ILLLLAVIL TYILLAMLI ALMGFTVNKI AQESKNIWL
701 QRAITILOTE KSEFKCHRA FRSGKLLQVG YTPDGKDYR WCFRVDENVN
751 TTNWNTVGI I NEDPGNCEGV KRTLSFLRS SRVSGRHWKN FALVPLLEA
801 SARDRQSAQ EEVYLRQFSG SLRPDAEVF KSPAASCEK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_20k2, frame 2

TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBL:AB015231.1 product: "stretch-inhibitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhibitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus Vanilloid receptor subtype 1 mRNA, complete cds.
Length = 838

HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query: 1 MKKWSSTDLGAAADPLQKDTCPDPLDGDPSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
M++ +S D + P Q++C DP D DPN +PPP KP + T +SRTRLFGKGDSEEA P
Sbjct: 1 MEQRASLDSEESPPQENSCLDPPDRDPNCKPPVPKPHITTSRTRLFGKGDSEEAFF 60

Query: 61 VDCPHEEGELDSCPTITVSPVITIQRPDGGTGARLLSQDSVAASEKTLRLYDRRSIFE 120
+DCP+EEG L SCP ITVS V+TIQRPDGGP R SQDSV+A EK RLYDRRSIF+
Sbjct: 61 LDCPYEEGLASCPITVSSVLTIQRPDGGPASPVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query: 121 AVAQNNCOOLESLLLFLQSKKHLTDNEFKDPETGKTKLLKAMLNHOGQNTTIPILLEI 180
AVAQ+NCQ+LESLL FLQ+SKK LTO+EFKDPETGKTKLLKAMLNH+GQN TI LLL+
Sbjct: 120 AVAQSNCOELESLLFLQSKKRLTDSEFKDPETGKTKLLKAMLNHNGQNTIALLDV 179

Query: 181 ARQTDLSKELVNASYTDSYKGTALHIAIERNNMALVTLVENGADVQAAANGDFFKKT 240
AR+TDSLK+ VNASYTDSYKGTALHIAIERNN LVTLLVENGADVQAA+GDFFKKT
Sbjct: 180 ARKTDLSKQFVNASYTDSYKGTALHIAIERNNMTLVTLLVENGADVQAAANGDFFKKT 239

Query: 241 KGRPGFYFGEPLPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNVTLHALVEADNTA 300
KGRPGFYFGEPLPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNVTLHALVEADNT
Sbjct: 240 KGRPGFYFGEPLPLSLAACTNQLAIVKFLQNSWQADISARDSVGNVTLHALVEADNTV 299

Query: 301 DNTKFTVSMYNEIILGAKLHPTLKLEELTNKKMGTPPLAACTGKIGVLAYILQREIQE 360
DNTKFTVSMYNEIILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct: 300 DNTKFTVSMYNEIILGAKLHPTLKLEELTNRKGLTPLAALSSGKIGVLAYILQREIHE 359

Query: 361 PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 420
PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN
Sbjct: 360 PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 419

Query: 421 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRVPDGLPPPFMEK-IGDYFRVTGEI 479
RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRVPDGLPPPFMEK+K++ +GDYFRVTGEI
Sbjct: 420 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRVPDGLPPPFMEKNTVGDYFRVTGEI 479

Query: 480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLVYFSLKEYVA 539
LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+OSLFML +VVLVYFSL KEYVA
Sbjct: 480 LSVSGGVYFFFRGIQYFLQRRPSLKSFLVDSYSEILFFVQSLFMLVSVVLVYFSQRKEYVA 539

Query: 540 SMVFSALGWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVYIVFLFGFSTAVVTLI 599
SMVFSLA+GWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVY+VFLFGFSTAVVTLI
Sbjct: 540 SMVFSLANGWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVYVFLFGFSTAVVTLI 599

Query: 600 EDGKNSLPSESTSHRWGPACRPDSSVNSLYSTCLELKFRTIGMGDLEFTEYDFKAV 659
EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCLELKFRTIGMGDLEFTEYDFKAV
Sbjct: 600 EDGKNSLPSESTPHKCRGSACKP-GNSYNSLYSTCLELKFRTIGMGDLEFTEYDFKAV 658

Query: 660 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSLKCMRK 719
FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSLKCMRK
Sbjct: 659 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSLKCMRK 718

Query: 720 AFRSGKLLQVGYTPDGKDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR 779
AFRSGKLLQVG+TPDGKDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR
Sbjct: 719 AFRSGKLLQVGYTPDGKDDYRWCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR 778

Query: 780 SSRVSGRWKKNFALVPLLRASARDQSAQPEEVYLRQFSGSLKPDAEVFKSPAASGEK 839
S RVSGR+WKNFALVPLLR+AS RDR + Q EEV L+ ++GSLKPDAEVFK GEK
Sbjct: 779 SGRVSGRWKKNFALVPLLRDASTDRHATQQEEVQLKHYTGSLLKPDAEVFKDSMVPGER 838
```

Pedant information for DKF2phtes3_20k2, frame 2

Report for DKF2phtes3_20k2.2

[LENGTH] 839
[MW] 94950.75
[PI] 6.90
[HOMOL.] TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus
vanilloid receptor subtype 1 mRNA, complete cds. 0.0
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05
[PIRKW] alternative splicing 3e-06
[PIRKW] peripheral membrane protein 3e-06
[SUPFAM] ankyrin repeat homology 3e-06
[SUPFAM] unassigned ankyrin repeat proteins 3e-06
[PFAM] Ank repeat
[KW] TRANSMEMBRANE 4

SEQ MKKWSSTDLGAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTRLFGRGDSEEAFF
PRD ccc
MEM

SEQ VDCPHEEGELDSCPTITVSPVITIQRPDGPPTGARLLSQDSVAATEKTLRLYDRRSIFE
PRD ccc
MEM

SEQ AVAQNQCQDLESLLLFLQSKKKHLTNEFKDPETGKTCLLKAMLNLDGQNTTIPLLLEI
PRD hhh
MEM

SEQ ARQDLSKELVNASYTDVYKGTALHIAIERRMALVTLLVENGADVQAAAHGDFFKKT
PRD hhh
MEM

SEQ KGRPGFYFGELPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNVTLHALVEADNTA
PRD ccc
MEM

SEQ DNTKFVTSYMEIILIGAKLHPTLKEELTNKKGMTPLAAGTGKIGVLAYILQREIQE
PRD chhh
MEM

SEQ PECRHLSRKFTWAGPVHSSLYDLSCIDTCEKNSVLEVIAYSSETPNRHDMLVEPLN
PRD ccc
MEM

SEQ RLLQDKWDRFVKRI FYFNFLVYCLYMIIFTMAAYRVDGLPPFKMEKIGDYFRVTGEIL
PRD hhh
MEM

SEQ SVLOGVYFFFRGQYFLQRRPSMKTLEVDYSSEMLFFLOSFLMATVVLVYFSLKEYVAS
PRD ccc
MEM

SEQ MVFSLALGWTNMLYYTRGFQMGIVAVMIEKMLRLDLRFMFVYIVFLFGFSTAVVTLE
PRD hhh
MEM

SEQ DGNKDSLPESTSHRWGFPACRPPDSSYNLSYSTCLELKFKTIGMGDLFTENYDFKAVF
PRD ccc
MEM

SEQ ITLLAYVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDEKSFLLKCMRKA
PRD hhh
MEM

SEQ FRSGKLLQVGYTPDGKDDYRWCFRVDEVNMTTWNINVGIIINEDPGNCEGVKRTLSFSLRS
PRD hhcc
MEM

SEQ SRVSGRWKKNFALVPLLREASARDQSAQPEEYILRQFSGSLKPEDAEVFKSPAASGEK
PRD ccc
MEM

(No Prosite data available for DKF2phtes3_20k2.2)

Pfam for DKF2phtes3_20k2.2

WO 01/12659

PCT/IB00/01496

HMM_NAME	Ank repeat
HMM	*GyTPLHIAARYNNVEMVc1LLQHGADIN*
	G+T+LHIA +++N+ +V LL+++GAD+
Query	202 GQTALHIAIERRNMALVTLLVENGADVQ 229

WO 01/12659

PCT/IB00/01496

DKF2phtes3_2013

group: transmembrane protein

DKF2phtes3_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp

Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCTCGGTT GTTGTGAGT GGAGACGAG GAGTGGGGCC
51 AGCCACGAGA AACAGTGGGC TGTACACAT CACCTTCAAA TATGACAATT
101 GTACCACCTA CTTGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCACG
151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCTGAAJ GATTTTCGGG
251 TAATACTGGA GGAGCTGAAG TCGGAGGGA GACAGTGCCT ACAACTGATT
301 CTAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCTTCAATA TGAAATTTGA AACGGATTAT TTGCTAAAGG
401 TTGTCCCTTT TCCTTCCATT AAAACGAAA GCAATTACCA CCCTTTCTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGGACA ATCTAGCTTG
501 TAAACCCCTC TGGAAGCCTC GGAACCTGAA CATCAGCAG CATGGCTCGG
551 ACATGCAGGT GTCTTCGAC CACGACCCGC ACACTTCGG CTTCGGTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
701 TTTCTCCAGG GATTATATA ATTGAGCTGG TGGATGACAC TACACACACA
751 AGAAAGTGA TCATTATGCT CTTAAAGCCA GTGCACTCCC CGTGGGCGGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGTAGTCT ATATCGGCAT
851 TCGCGACGCT CTTCACTGTG ATGTGCGCGA AGAAGCAACA AGAAAATATA
901 TATTACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT AACTGCAGC
951 ACTCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGTCTTT CTCTGCTATT
1001 CCAGTAAGA TGCCGAGAAT CACATGAATG TCGTCCAGTG TTTCGCTTAC
1051 TTCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTGAGCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTCTAT CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGT CCAAGGTAT GAAGTACTT
1201 GTGGCAAGA AGAATACAA ACACAAGGA GTGTGCGGAG CTTGGGGGAA
1251 AGGAGAGCTC TTCTTGGTGG CGGTGTGAGC CATTCGCGAA AGCTCCGCC
1301 AGGCCAAGCA GAGTTGCTCC GCGCGCTCA GCAAGTTTAT CGCGTCTAC
1351 TTTGATTATT CTTGCGAGGG AGACGTCCCC GGTATCTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCTCA GCTCTGTTCC CACCTGCATC
1451 CCCGAGACCA CGGCTCCAG GAGCCGGGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAAC ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACGTCGCCAT
1551 TTGCAACATG CACCAAGTTA TTGACGAGGA GCCCGACTGG TTGGAAGAGC
1601 AGTTCTGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTTG ATTCCGGCTT GGTTTTAAAT GATGTCATGT CCAACCGGG
1701 GCTGAGAGT GACTTCTGCT TAAAGTAGA GCGGCTGTT TTTGGGGGCA
1751 CCGGACGAGC CGACTCCAG CACGAGAGTC AGCATGGGGG CTTGGACCAA
1801 GACGGGGAGG CCGGCGCTGC CTTGACGGT AGCGCCGCC CTTGAAACCTC
1851 GCTGCACAGG GTGAAAGCGG GCAGCCCTC GGACATGCGG CGGAGCTCAG
1901 GCATCTATGA CTCGTCTGTG CCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCTCTCTCT TCAGGCTGG GTGAGGAGGA ACCTCCTGCC CTTCTTCCA
2051 AGCTCCTCTC TTCTGGGTCA TGCAAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGCGGT CGCCCTTTTG TAAACAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTAGCTGC TGCCCTCCCTC TGATTCCCA GCTCATCTCC
2201 CTGGTTGCAT GGCCCACTTG GAGCTGAGGT CTATACAG GATATTGGA
2251 GTGAATGCT GCGCACTACT TGTCTCCCT TGCCCAAC CTTTACCGGA
2301 TATCTTGACA AACTCTCAA TTTTCTAAA TGATATGAG CTCTGAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2401 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 346 bp to 2130 bp: peptide length: 595
Category: similarity to known protein
Classification: unclassified

```
1 MESQPFLNMK FETDYFVKV PPSIKNESN YHPFFFRTRA COLLQPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKHGEGPKR
101 KTKQEQTTE MTSCLLQNVN PGDYIIELV DNTTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAF A TLFTVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNHMNVV QCFATFLQDF CGCEVALDLW
251 EDLSLCREGQ REWVIOKHE SQTIVVCSK QHRYFDKKN YKHGGGRGS
301 GKGLFLVAV SAIAEKLRQA KQSSAALS K FIAVYFDYSC EGVVPGIIDL
351 STKYLMDNL POLCSHLHSR DGLQEPQGH TROGSRNRYF RSKSGRLYV
401 AICNMHQFID EEPDFWEKQF VPFNPPPLRY REPVLKFDG GLVLNDVMCK
451 PGFESDFCLK VEAALVLTATG PADSQHSQH GGLDQDGEAR PALDGSAAQ
501 PLLHTVKAGS PSDNPRDSGI YDSSVPSEEL SLPLMEGLST DQETSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGCKADL GCRSYTDELH AVAPL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2013, frame 1

TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus
interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =
1.1e-13

>TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
mRNA, complete cds.
Length = 866

HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14
Identities = 85/284 (29%), Positives = 131/284 (46%)

```
Query: 213 KVFLEYSSKDGQNHMNVVQCFAFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
      KV++ YS+ D +++VV FA FL CG EVALDL E+ ++ G WV QK +
Sbjct: 379 KVMIIYSA-DHPLYVDVVLKFAQLLTACGTEVALDLLEEQAISEAGVMTWVGKQKQEMV 437

Query: 269 HESQFIIVVCSKGMKY----FVDKKNXXXXXXXXXXKELFLVAVSAIAEXXXXXXXX 324
      + IIV+CS+G + + + +LF A++ 1
Sbjct: 438 ESNKIIIVLCSRGTRAKWQALLGRGAPVRLRCDHGKPVGDLTAAAMNMLPDFKRPACFG 497

Query: 325 XXXXXXFIAYVF-DYSCGQDVPGLDLSTKYRLMDNLPLCSHLHSRDHGLQEPGQHTRO 383
      ++ YF + SC+GDVP + + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGDVPLFGAAPRYPLMDRFEEV--YFRIQDLEMFPQGRMHRV 550

Query: 384 G--SRNRYFRSKSGRSLYVAICNMHQFIDEEPWFKEQFV----PFNPPPLR---YREP 434
      G S NY RS GR L A+ + PDWFE + + P L + EP+
Sbjct: 551 GELSGDNYLRSPGGRLRAALDRFRDQVRCQDWFECENLYSADDQDAFSLDEEVFEPL 610

Query: 435 LEKFDSGLVLNDVMCKPGPESDFCLKVEAAVLGATGPADSQHSQHGLDQDGEARP 491
      L +G+V + + P S CL ++ V G GA H L G+ P
Sbjct: 611 LPP-GTIVKRAPLVRE-PGSACLAIDPLV-GEEGAATAKLEPH--LQPRGQAP 662
```

Pedant information for DKFZphtes3_2013, frame 1

Report for DKFZphtes3_2013.1

```

2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAATGC
2651 AGTGCTGACA TCAGTGAGTT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAACTGTAGC CTTCATGGGC
2801 ACGCCTCTGC TGGAAACGTC CAGCACGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCCTGT TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAGTGAT TCTTGTGGAA AATGAGAGGT GAGCTCATTC TTCTGAAATG
2951 GTCCCCCTAT CCTGGAAATC AGTGGGGAGA GGTTTTTGAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCACT TGGGGACCAA
3051 ACAAGAGACC CTCCGCATTG CAGCCTGCAG TTGCCGCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAAATGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAATCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCAAGT GATAGCACCT AATCAATTCC TCACACGTCA AGTGACACCT
3301 CAGACATGGG ATAGATTTCC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAGCA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCAGC CTTTTCAGGC ACGGAGATTG CCTTTCAACA
3451 TCCAAACATT TCCCAGAAC CATGTGCCAT CCTACTTGTA TACTGGTGG
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCT ATTTTATTTC
3551 ACGAGAACAG CACATACATG TGTTTGAAAA TTATGTGAGG TGCTCACTCT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCTGCCCC ACCTTGACAG
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCC
3701 TGGAGATCAC ATCTTCCATT GTTCCTTCAA TCAACACTAA CTCCCATTG
3751 GGCCTTAGGT GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CCAGCCCGCT CCTACTGAGG GCCTCCTCTC TGTGAGGCAC
3851 CTTGCAAAGC ATTTTGTGTG AAGTGACTCA TTTAACCTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 GCCCAGAGAG GTGAAATGCC TTCCCGAGAG TTACACAGAG TGCTGGAGCT
4001 GGAATACTG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TGCTACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACAGG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTTACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAACTGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAAACA TGTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTAT CGAAAAATGC CAAAGCCTAT ATCCTTTCTG TGATGACCTT
4501 CCCCATGGGG AGGTGCTACA GAGCCCTGG GCTTGTCCTG GCCTCTGGAC
4551 AAAAGAATGT TCCACAGGGT CTGAGGAGGT TTCCCGACCC TCAGAACAAT
4601 GATGGCCTGG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ACTGTTTGCA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTT
4701 ACTAGGAAAT GACGCCCCCT GTGTTGCAGG CAAGCACACT CTGGGGTTGA
4751 GGCAACCCCC ACGTGGAAGA CACTATAAGG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCCT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGCCAGGAG CCTCCATGCA GGAGGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGCCC
5101 AGGGTTCGGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TTGCCTTGGG CCTTAAAGT CTGTCTCCCG CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGG CTGCTCCTCC CGGCGCTTGG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCCCTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCTTAAC GACCTGCGCG CGCTTTTGT CGATAAAGAT ACGATTGTTA
5401 ATGCTGTCGG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCAG AATCAACTCT TGGGTACAC GTTTAATAGA
5501 CAGGATTAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCACATCC TAGACTAGAT GAATGTCAGC CACAGGAGCT TCTTCAAAAC
5651 ATAGCACCA GAGAGGAAAG GAGAACGCA TGCACACGCG TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATCT
5751 TCCCCACCC CTGGAAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HSI292248 from database EMBL:

human STS SHGC-53917.

Score = 874, P = 3.3e-33, identities = 180/185

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225
Category: similarity to known protein

```

1 MNQPCNSMEP RVMDDDLKLL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL
51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGENLAHLV WLDLSFNNIE
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
151 YLRRFKCLRT LSLSRNPIS EADYKMFICA YLPDLMYLDY RRIDDHTASV
201 SLSVSQPCET DSSSPQVSWK RGIEE

```

BLASTP hits

Entry S68209 from database PIR:
sds22 protein homolog - human >TREMBL:HSSDS22MR.1 gene: "sds22";
product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA
Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:
suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)
>TREMBL:SPSDS22.1 gene: "sds22+"; S.pombe sds22+ gene, complete cds.
Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:
protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)
>SWISSPROT:SD22 SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT
SDS22. >TREMBL:SPAC4A8.12 gene: "sds22"; product: "phosphatases pp1
regulatory subunit"; S.pombe chromosome I cosmid c4A8.
Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2.5 from database TREMBL:
gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.
Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

Alert BLASTP hits for DKFZphut1_20ml1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_20ml1, frame 1

Report for DKFZphut1_20ml1.1

```

[LENGTH]      225
[MW]           25955.87
[pI]           4.63
[HOMOL]        PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]       30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
[FUNCAT]       01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
YJL005w] 3e-05
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]       04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
[EC]           4.6.1.1 Adenylate cyclase 2e-06
[PIRKW]        nucleus 5e-16
[PIRKW]        duplication 2e-06
[PIRKW]        tandem repeat 2e-06
[PIRKW]        cAMP biosynthesis 2e-06
[PIRKW]        glycoprotein 2e-06
[PIRKW]        phosphorus-oxygen lyase 2e-06
[SUPFAM]       leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM]       fibromodulin 3e-07
[SUPFAM]       yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM]       yeast adenylate cyclase 2e-06
[PROSITE]      CK2_PHOSPHO_SITE 2
[PROSITE]      PKC_PHOSPHO_SITE 1

```

[KW] All_Alpha

```
SEQ  MNQPCNSMEPRVMDDDMLKLAVGDQGPQEEAGQLAKQEGILFKDVLSLQLDFRNILRIDN
PRD  cccccccccccccchhhhhccccchhhhhhhhhhhchhhhhhhhhcccccccccccc

SEQ  LWQFENLRKLQLDNNIIEKIEGLENLAHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLVLSLGNNRIDNMMNIIYLRRFKCLRTLSLSRNPISAEADYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccccccchhhhhhhhhccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVLSVSPQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccc
```

Prosite for DKFZphutel_20ml1.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphutel_20ml1.1)

DKFZphute1_20m24

group: metabolism

DKFZphute1_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits
Alg9 is involved in the assembly of the core oligosaccharide
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)

HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```

1  TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51  TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCGGCG TCGGGACAAG
101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GGCGGCGCGG AGCACC GGAC
151 CGAGTTATCT GGGAAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGTCTGCTT TCAGCAAGGT TATGTGCTGC TCTCCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTCGCTC CTATGCTTAC CTGTGTGCTT ATGCCTGGCC AGCTGCATTT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTTGTGTTTT ACTTTTTCG
451 ATGTCTTCTG GCTTTTGTGA GCTGTATTG TGAACCTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATTGCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTGG ATTTGCTGGT
751 CATGAAACAC AGGTGGAAGA GTTTCTTTCA TTGGTTCGCTG ATGGCCCTCA
801 TACTATTCTT GGTGCCTGTG GTGGTCATTG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTA CTCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCCTG GTATTCTCTAT TTAATTAATG
951 GATTCTTGAA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCTTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAAATT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTTCTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCCTTT CCCTGTGTAT
1151 CCACTTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTTGTCATTT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTTGTA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGTA TCGATTTCCT
1451 AGCAGCTTCC TTCTTCCTGA CAATTGGCAG CTTCAGTTCA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGTTG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTCATC CAATAAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTC CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCCTT CCTGTCAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAACCCC GGAAAGCAAA GCAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSAC381 from database EMBL:
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.
Length = 601

Medline entries

96293493:
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: identification of the ALG9 gene encoding a putative mannosyl transferase.

Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611
Category: strong similarity to known protein

```

1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTELSGNKA
51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEFG
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVEY FLRCLLAFVS
151 CICELYFFYKA VCKKFGLHVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPF S AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVEL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSF L LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVLDL TMRETPREP K
551 YSSNKEEWS LAYRPF LDAS RSSKLLRAFY VFFLSDQYTV YVNYTILKPR
601 KAKQIRKKSG G

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20m24, frame 2

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.

Length = 653

HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query:  48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEP THYLIYGEFGQTWEYSP 107
        N   W   + FK LLS R+ A+ I+DCDE +NYWEP H +YGEFGQTWEYSP
Sbjct:  43 NNPNDNWPFSSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFGQTWEYSP 102

Query:  108 AYAIRSYAYLLHAWPAAFHARILQTNKILVEYFLRCLLAFVSCICELYFYKAVCKKFGL 167
        YAIRSY Y+ LH PA+ A+ KI+VF +R + + E Y + A+CKK +
Sbjct:  103 VYAIRSYFYIYLHYIPASLFANLFGDTKIVFTLIRLTIGLFCLLGEYAFDAICKKINI 162

Query:  168 HVSRMLLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
        R + F + S+GMF +S+AF+PSSFCM T + + + + + VA ++GW
Sbjct:  163 ATGRFFILFSIFSSGMFLASTAFVPSFSCMAITFYILGAYLNENWTAGIFCVAFSTMVWG 222

Query:  228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```

Sbjct: 223 PFSA LGLPI D+L++K F SL+ + V+ DS+Y+GK V+APLNI LY
 Query: 288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFA LALLVPLTSLMEYLLQRFHVQNLGHPY 347
 NV + GP LYG EP FY+ N F N+N+ A PL+ + Y + + Q+
 Sbjct: 283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340
 Query: 348 WLT LAPMYI-----WFIIFFIQPHKEERFLFPVYPLICLCAVALSALQKCYHVFQR 400
 + AP+ + W +IF Q HKEERFLFP+YP I A+AL A + ++
 Sbjct: 341 YQRFAPIIILLAVTTAAWLLIFGSAHKEERFLFPYPIAFFAALALDATNR---LCLKK 397
 Query: 401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPE 460
 ++ N L++ + F +LS SR+ ++ Y +++Y T+ T +
 Sbjct: 398 LGMD-----NLSILFILCFAILSASRTYSIHNNYGSHEIYRSLNAELTNRT-NFRNF 450
 Query: 461 GRPVNVCVGKEWYRFPSSFLPDNW-----QLQFIPSEFRGQLPKPFAEGPL---ATRI 511
 P+ VCVGKEW+RFPSSF +P +++FI SEFRG LPKPF + TR
 Sbjct: 451 HDPIRVCVGKEWHRFPSSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPFLLKSDKLVEVTRH 510
 Query: 512 VPTDMNDQNLLEEPSRYIDISKCHYLVDLDTMRETREPKEYSSNKEEW 558
 +PT+MN+ N EE SRY+D+ C Y+VD+D M ++ REP + ++ +
 Sbjct: 511 IPTEMNNLNQEEISRYVDLSDCDYVVDVD-MPQSDREPDFRKMRQNY 556

Pedant information for DKFZphutel_20m24, frame 2

Report for DKFZphutel_20m24.2

[LENGTH] 611
 [MW] 69863.78
 [pI] 8.91
 [HOMOL] SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-93
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69
 [PIRKW] glycosyltransferase 9e-68
 [PIRKW] transmembrane protein 9e-68
 [PIRKW] hexosyltransferase 9e-68
 [PROSITE] MYRISTYL 9
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] PKC_PHOSPHO_SITE 6
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] TRANSMEMBRANE 7
 [KW] LOW_COMPLEXITY 6.71 %

SEQ MASRGARQRLKSGASSGDTAPAADKLELLGSREAGGAEHRTLSGNKAGQVWAPEGST
 SEG
 PRD cchhhhhhhccccccccccccchhhhhhhhhccccccccceeeccccccccccccch
 MEMMMMMMM

SEQ AFKCLLSARLCAALLSNISDCDEFNYWEPHYLIYGEFQTWEYSPAYAIRSYAYLLLH
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhccccceeeccccceeeccccceeeccccchhhhhhhhhhhc
 MEM MMMMMMMMMMMMMMMMMM.....M

SEQ AWPAAFHARILQTNKILVIFYFLRCLLAFVSCICELYFYKAVCKKFLHVSRLMLAFLVLS
 SEG
 PRD cchhhc
 MEM MMMMMMMMMMMMMMMMMM.....

SEQ TGMFCSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD
 SEG
 PRD cceeeccccccccchhh
 MEMMMMMMMMMMMMMMM

SEQ LLVMKRWKSFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHGPDLYGT
 SEG
 PRD hhhc
 MEM MMMMMM.MMMMMMMMMMMMMMMMMM.....

SEQ EPWYFYLINGFLNFNVAFA LALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII
 SEG
 PRD cceeeccccccccchhh
 MEMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

Prosite for DKFZphutel_20m24.2

(No Pfam data available for DKFZphutel_20m24.2)

DKFZphutel_21d15

group: uterus derived

DKFZphutel_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGCGGGCCT GCGGGCCGTT
251 GTCCGGGGCA CTGCGCAGCG CGGGCACCCC CGCGGCCCTT CCCCTGGGCG
301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GCGCGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCCTCGA GGCCGCAGTG GCGCGCGCCA
401 GACCACTGCC AGGCTCAGGC GCGGCGCGGG CTGGGCGACG GCGAGGACGC
451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTT GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCGG
601 GCCCCAACAC CGTGCCCAAG GCCCGGGGCC CGAGCCCACC TTAGGTCCCC
651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
701 CAGGCCCTTG CACGGCTGAG CCCGAGAGGA CCGCGCTCA GCCCGGGTCC
751 CCACGCTGCC CCCGCGCTG CTCTGCGTCG GTCCGCGCGG CTCCCACTCA
801 CTCGCTGCTG GTCGCTCTCC GGGCGGGGGC GACTTGGCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCGGGCTT CCTGACTGAT
1001 GAAGAGTGTG GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTACAGCG
1051 CAGCCAGATC CTGCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTGAGCCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCGTG
1301 AGCCTCTCTG CCACTCCAG GTGCACAATT TTGAAAACCT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTG GTGATGGTCT CGAGGGCAGT TCTTGGAGAC
1501 CTTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCTT GCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCCT TCCTCTTGTC CACCTGGAGT CATGCCGAAG CGCCTAAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTGATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGG AGGTGATGTC CTTGGGTGAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTCTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTTGC TAAATGGGCT GGTAAACGAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCCC AGGATGAAAC CTGGGGTCAT GAGGAACTCC
2301 CCGGGGGCTG GCCCTGCTTG CACCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGGCCATATG
2451 GCTCAAGCTT CTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAACCTGAG AAAAGGAACT
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTG CAAGGCCCTT CAGTGCCTGA
2651 GCTCAGGGGC TGCTGCTGCC TTTGGGTTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCT CCCAATATGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTGGGGAGT
2801 TCCACAAGTA CATGAGGAGC CACAAGGAGC AGTCCAGTGA GCTGGTGCGG
```

```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGGCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGTCCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCCGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCCTCTGT CCCACCAGGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCAGCCG CTGCAGGTG TTCGATATGG TGAGGGGGGC
3601 CACTACCATG CCCACGTGGA CAGTGGGGCT GTGTACCCAG AGACCATCTG
3651 CTCCCATACC AAGCTGGTAG CCAACGAGTC TGTACCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCCACAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGGTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAACTTG
3951 CCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCCTA GCTTGTCTG
4001 CCCATTCTCT CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCTCTCCTC CCCTGTCAAG CCAGTCTCTT CCTCTTCAGG TGGCTGTTCT
4101 GGCCAGCCCC CTTCCCATCC CCAAGGAGCC CTTCAGCGCG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCC TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CTTGGGCAGC TTATCTTGCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCTT ACAGCTACAT GACAGTGCTG
4451 TTTTATTGTA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTTTGGGCTG GCAGGGGCTT AGACAAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTACCCC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCAACCT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGAAA CCTGCGTGTC AAGCCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACCTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGATTGCCA ACAACTGCAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCAA CAGGAGATGG CCCGCTTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCCA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAGGTCTT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCTTAT ATTCAATGTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAGG GG

```

BLAST Results

Entry HSU64252 from database EMBL:
 Human STS sequence NOTI-225.
 Score = 959, P = 1.2e-36, identities = 195/199

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118
 Category: questionable ORF
 Classification: no clue

```

1 LPLVYALMVP LLSASTLGLT ASDLESVQLC PTATQLGKRS PSVGWGSRRR
51 KAEPEGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGDRPAA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 1

No Alert BLASTP hits found

Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191

Category: putative protein

Classification: no clue

1 MAAAVTQQR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAAARCAFF P

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1
Length = 1,298

HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 36/103 (34%), Positives = 44/103 (42%)

Query: 87 GDESSDPGPQHRAQGGPGPEPTLGLPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
G + PGP G GP P P T+ G S R P PA S P GP +P
Sbjct: 726 GRKRKSPGPARPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARS GAPAAARCAP 189
AAP AA ++R P+ GP LG W + P+ AP
Sbjct: 783 PAAPRAAAQARPRPVAVSRRPAEGPDPLGG-WRRQPPGPSHTAAP 827

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 8/21 (38%), Positives = 9/21 (42%)

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48
DH + A G G AP P
Sbjct: 212 DHAREARAVGRGPSSAAPAAP 232

Pedant information for DKFZphut1_21d15, frame 1

Report for DKFZphut1_21d15.1

[LENGTH] 117
[MW] 11797.32
[pI] 10.68
[KW] Irregular
[KW] SIGNAL_PEPTIDE 22
[KW] LOW_COMPLEXITY 38.46 %

SEQ LPLVYALMVPLLSASTLGLTLASDLESVQLCPTATQLGKRSPSVGWSRRRKAEPGADAGG
SEGxx
PRD cccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccc

SEQ SGRAQHPQAPSPSDRGARGPGGRC PGDCAARAPPRLPWARARPGCHGGSGGDRPAA
SEGxx
PRD ccc

(No Prosite data available for DKFZphut1_21d15.1)

(No Pfam data available for DKFZphute1_21d15.1)

Pedant information for DKFZphute1_21d15, frame 2

Report for DKFZphute1_21d15.2

```
{LENGTH}      191
{MW}           19916.88
{pI}           10.43
{KW}           TRANSMEMBRANE 1
{KW}           LOW_COMPLEXITY 29.84 %

SEQ  MAAAAVTGQRPETAAEEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG  .....
PRD  cccceeeccccchhhhhhhhhccccccchhhhhhhhhccccccccccccccccccccchhh
MEM  .....

SEQ  FLVLMVFVHLYLGNVLALLLVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccceeeee
MEM  ....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ  TSLPRRAPGPARLSPRGPALSPGPHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eccccccccccccccccccccccccccccccccchhhhhhhhhccccceccccccccchhhhhc
MEM  .....

SEQ  GAPAAARCAPP
SEG  xxxxxxxxxx..
PRD  cccccccccc
MEM  .....
```

(No Prosite data available for DKFZphute1_21d15.2)

(No Pfam data available for DKFZphute1_21d15.2)

DKF2phutel_22d2

group: signal transduction

DKF2phutel 22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp
Poly A stretch at pos. 3230, no polyadenylation signal found

```

1 CTCCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
151 CCAGAAAGAGG TTCCTCCCCG GGCAGAAAGAA ATCACCATTG CAGCTGATGT
201 CACCCAGAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCATCTGT
301 ATAGTGATAT CCGTTAAACA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAATATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTACGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGACTGCCC AGAGGAGAAG
601 GAGATGAAGC CAGCTTGATG AAAAGCCCTT ACTCGTATAT TTAATAATATC
651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAACTC AACTTCTTTC
701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAGCTCTT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCCGTGAA GGTTTTCTCT TTTTACACAC ACTTTTATC CAGAGAGGGA
851 GACAGGAAAC TACTTGGACT GTGCTTCGAC GATTGGTTA TGATGATGAC
901 CTGGATTGGT CACCTGAATA TTTGTTCCSC CTGCTGAAA TACCTCTCTGA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTGGGAT AGAGACTGTG CTTTGTCAAC TGATGAGCTT
1051 AAAGATTATG TTAAGTTTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCCAGTG GACGCTCAGC ACTTATTTAG ATGTACAGCG GTGCTTGGA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTCAAGTGTG ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAATCTCAAG AAATGTGTTT AGATGTAATG TAATTGGAGT GAAAAACTGT
1351 GGGAAAAGTG GAGTTCTTCA GGCTCTTCTT GGAAGAAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTGT
1451 ATGTATATGG ACAAGAGAAA TACTTGTGTT TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTTGT GATGTTGTAT GCCTGGTATA
1551 TGATGTGAGC AATCCCAAAT CCTTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTACCTTA CTGATTCTTG
1701 CAGGAAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGCAATGGT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAATT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGTAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCCAC CCAGCATTA AATAAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTTT
2201 CTAATAATTT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCACTTAG
2251 GATATGTCTT TTTTAAGTGC TGTAAGAGAG AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTTACG AGCCACAAAT TTCATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAATTTG CATGGTCTT AATGGCATTG
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTCT TTTTGTAACA AATAAGTTTT
2501 TCAGGCAAGT CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG

```

```

2601 AGTAATAATC AAACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTGTGATTG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTTGTAT
2751 TCGAAGACAA GTCATTCCCTA TTATTATAGA ATAACCAAAA CCTTATTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTTCTGTGGG ATTTTGTGA TATTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTTA ATCTTGTGAT TACAGTAACT GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAATTCTG TTGTGTGTGT
3201 GTGTGTTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAA

```

BLAST Results

Entry AC004527 from database EMBL:
 *** SEQUENCING IN PROGRESS ***. NF1-related locus, Direct Submission;
 HTGS phase 1, 10 unordered pieces.
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HSI48355 from database EMBL:
 human STS SHGC-31220.
 Score = 1826, P = 7.5e-78, identities = 388/406

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 64 bp to 1803 bp; peptide length: 580
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSI IMSLVSEEF EEVPPRAEEI TIPADVTPER
51 VPTHIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPIMNQYTE IETCVCESAK
151 NLKNISELFY YAKKAVLHPT GLYCPEEKE MKPACIKALT RIFKISDQDN
201 DGTLNDAELN FFQRICFNTP LAPQALDVK NVVRKHISDG VADSGTLKLG
251 FLFLHTLFQI RGRHETTWTV LRRFGYDDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHLDR DCALSPDELK DLFKVFPYIP WGPDVNNTVC
351 TNERGWITYQ GFLSQWLTIT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDIQ KKQTQRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVYVYGQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFECYARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPQAFCTN TADAPSKDIF VKLTTMAMY

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22d2, frame 1

TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320_4 gene: "SPCC320.04c"; product: "hypothetical protein";
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11.

Length = 625

HSPs:

Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138
Identities = 263/582 (45%), Positives = 380/582 (65%)

Query: 4 DVRILLVGEPRVGKTSLSIMSLVSEEFPEEVPPRAEITIPADVTPERVPTHTIVDYSEAEQ 63
DVRIL+L+G+ GKTSLS+MSL+ +E+ + VP R + + IPADVTPE V T IVD S E+
Sbjct: 9 DVRIVLIGDEGCGKTSLSVMSLLEDEWVDAVPRRLDRVLIPADVTPENVTTISIVOLSIKEE 68

Query: 64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
+ + EI QANVIC+VY+V ++ +D + ++W+PLI + + P+ILVGNKSD
Sbjct: 69 DENWIVSEIRQANVICVVSVDDESTVDGIQTKWLPILIRQSFGEYHETPVILVGNKSDGT 128

Query: 124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYQAQKAVLHPTGPLYCPEEKEMKP 183
++ + ILPIM TE+ETCVECSA+ +KN+SE+FYQAQKAV++PT PLY + K++
Sbjct: 129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYQAQKAVIYPTRPDYDADTKQLTD 187

Query: 184 ACIKALTRIFKISDQNDGTLNDAELNFFQRICFNTPLAPQAEDEVKNVVRKHISDGVAD 243
KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+
Sbjct: 188 RARKALIRVKICDRDNDGYLSDELNDQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query: 244 SGLTLKGFLFLHTLFIQRGRHETTWTVLRRFGYDDDLDTPEYLFPLLKIPPDCTTELNH 303
L L GFL+LH LFI+RGRHETT WLR+FGY+ L L+ +YL+P + IP C+TEL+
Sbjct: 248 DSLMLAGFLYLHLLFIERGRHETT WVLRRKFGYETSLKLSYDLYPRITIPVGCSTELSP 307

Query: 304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQGFL 363
F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++
Sbjct: 308 EGVQFVSALFEKYDEDKGCLSPSELQNLFSVCPVPVITKDNILAETNQRGWLYTYNGYM 367

Query: 364 SQWTLTLYDVQRCLLEYLGYLSILTEQESQAS----AVTVTRDKKIDLQKKQTQRNVF 419
+ W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF
Sbjct: 368 AYNMNTTLINLTQTLEQLAYLGFPVGRSGPGRAGNTLDSIRVTRERKKDLNHGTDKRVF 427

Query: 420 RCNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLDHI--- 476
+C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLL ++
Sbjct: 428 QCLVVGAADAGKTVMQSLAGRMADVAQIGRRH-SPFVINRVRVKESKYLLLRVDVL 486

Query: 477 SESEFLTEAEIICDVVCLVYDVSNPKSFHEYCARIFKQHFMDSRIPCLIVAASDLHEVKQ 536
S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q
Sbjct: 487 SPQDALGSGETSADVAVFLYDISNPDSFAFCATVYQKYFYRTKTPCVMIATKVEREEVDQ 546

Query: 537 EYSISPTDFCRKHKMPPPPQAFCTNTADAPSKDIFVKLTMMAMP 580
+ + P +FCR+ ++P P F+ S IF +L MA+YP
Sbjct: 547 RWEVPPPEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

Pedant information for DKFZphutel_22d2, frame 1

Report for DKFZphutel_22d2.1

[LENGTH]	580
[MW]	66541.61
[pI]	5.56
[HOMOL]	TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT]	99 unclassified protcins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT]	03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT]	03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]	11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]	10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07

```

[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
1e-07
[FUNCAT]      08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT]      10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT]      09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 9e-04
[BLOCKS]      BL00410A Dynamin family proteins
[SCOP]         dlplk_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 2e-42
[SCOP]         dlguua_ 3.25.1.3.10 Rap1A [Human (Homo sapiens) 5e-59
[PIRKW]        transmembrane protein 1e-79
[PIRKW]        membrane trafficking 2e-06
[PIRKW]        acetylated amino end 3e-09
[PIRKW]        prenylated cysteine 3e-09
[PIRKW]        signal transduction 1e-07
[PIRKW]        transforming protein 3e-09
[PIRKW]        immediate-early protein 8e-06
[PIRKW]        alternative splicing 4e-08
[PIRKW]        P-loop 1e-10
[PIRKW]        lipoprotein 7e-10
[PIRKW]        proto-oncogene 3e-09
[PIRKW]        methylated carboxyl end 3e-09
[PIRKW]        membrane protein 3e-09
[PIRKW]        GTP binding 1e-10
[PIRKW]        thiolester bond 7e-10
[SUPFAM]       ras transforming protein 1e-10
[PROSITE]      ATP_GTP_A 2
[PROSITE]      MYRISTYL 3
[PROSITE]      EF_HAND 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 14
[PROSITE]      TYR_PHOSPHO_SITE 4
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         Ras family (contains ATP/GTP binding P-loop)
[KW]           Irregular
[KW]           3D

```

```

SEQ      MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEETIPADVTPEVPVTHIVDYSE
1jai-    ...EEEEEEETTTTCHHHHHHHHHHCCCCCCCCCEEEEEETEEEEEEEEEECCC

SEQ      AEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLINERTOKDSRLPLILVGNKS
1jai-    CGGGHHHHHHHHHHHTTEEEETTTTHHHHHHHH-HHHHHHHHHHHCTTT-TCEEEETTT

SEQ      DLVEYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQAVLHPTGLYCPPEEKE
1jai-    TTTTTTTTHHHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....

SEQ      MKPACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALEDVKNVVRKHISDG
1jai-    .....

SEQ      VADSGTLTKGFLFLHTLFTIQRGRHETWTVLRRFGYDDDLTPEYLFPLLKIPDCTTE
1jai-    .....

SEQ      LNHAYLFLQSTFDKHDLDLDRDCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQ
1jai-    .....

SEQ      GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
1jai-    .....

SEQ      CNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLHDISESE
1jai-    .....

SEQ      FLTEAEIICDVVCLVDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-    .....

SEQ      SPTDFCRKHKMPPOAFTCNTADAPSKDIFVKLTMMAMP
1jai-    .....

```

Prosites for DKFZphut1_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

Pfam for DKF2phute1_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK		
	++L+G+ V GK++L ++ EF+EE +P ++ T ++ ++		
Query	6	RILLVGEPRVGKTSLIMSLVSEEFPEE-VPPR-AEEITIPADVTPERV	52
HMM	LQIWDTAGQERYRSMRPYYRGAMGFMVYDITNRqSFENIr.NWweEIr		
	I D E+ + + +A+++ +VY+++N+ S ++++ +W++ I+		
Query	53	THIVDYSEAEQSDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVStEEGqFAREWGAIPFMETSAKT		
	+ D+D+ P +LVGNK+DL + ++T + +E+SAK+		
Query	103	ERTDKDSRLPLILVGNKSOLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
	N+ E F+ + +++L + . +++ +++++ + C+		
Query	152	LKNISELFYYAQKAVLHPT-----GLYCPEEKEMK-PACI--	186

DKFZphute1_22e12

group: signal transduction

DKFZphute1_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits
cornichon is required for signal transduction in the EGF-receptor
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp

Poly A stretch at pos. 499, no polyadenylation signal found

```

1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTCGCGGC
101 TCATCTTCTC CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACCTGGT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC ACTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTTGGAA
451 CAAAAAATA TTTTGCTGT ATTTTACCA TATAAAGTAT TAAAAAACA
501 TGAATAAATA AAAAAAATA

```

BLAST Results

No BLAST result

Medline entries

95300228:
cornichon and the EGF receptor signaling process are necessary for both
anterior-posterior
and dorsal-ventral pattern formation in Drosophila.

Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92
Category: strong similarity to known protein

```

1  MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKWWIP
51  ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398.1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
Length = 138

HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWIPELIGHTIVTV 60
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++
Sbjct: 1 MGAWLFILAVVNCINLFGQVHFTILYADLEADYINFIELCSKVNLITPEAALHGALS 60

Query: 61 LLLMSLHWFIPLLNPVATWNIYRM 85
L L++ +WF+FLNLPV +N+ ++
Sbjct: 61 LFLNGYWFVFLNLPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90
+YRMI+ALI
Sbjct: 123 LYRMIMALI 131

Pedant information for DKFZphut1_22e12, frame 1

Report for DKFZphut1_22e12.1

[LENGTH] 92
[MW] 10614.98
[pI] 5.04
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
5e-14
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGL054c]
2e-15
[PIRKW] transmembrane protein 2e-11
[PROSITE] CK2_PHOSPHO_SITE 3
[KW] SIGNAL PEPTIDE 33
[KW] TRANSMEMBRANE 2

SEQ MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWIPELIGHTIVTV
PRD ccchhhhhhhhhhhhhhhhhheeeccccccccccccccccceehhhhhhhhhhh
MEMMMMMMMMMMM

SEQ LLLMSLHWFIPLLNPVATWNIYRMILALIND
PRD hhhhhhhheeeccccchhhhhhhhhhhccc
MEM MMMMMMMMMMMMMMMMM..MMMMMM....

Prosite for DKFZphut1_22e12.1

PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1_22e12.1)

DKFZphut1_22n2

group: uterus derived

DKFZphut1_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```
1 ACAACAGGCT GGTGCTTGG CGTGAATCC TAAAGTGGCC TGGCTTIGAG
51 ACTGGAGTGA GACCCAGGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTGAAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAAC GACTCCGAAG AGTTGCAAG TATTTATACT
201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAACTG ATTCTGATTC
401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTTG CCAGTTTCTG CTGAAATTAA GGAACCTTTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAACT
551 GAAGCCTTTC ATTCTGATT TTATCCCAGC TGTGGGGGAT ATTGATGCAT
601 TCTTAAAGGT CCCACGTCTT GATGAAAAGC CTGACAACCT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCACGG TGCTCTCACT
701 CTGGTTAACA GAGAATTCTA AGCAGACAAA CATCACAAA CATATGAAAG
751 TAAAAGCCTT AGAAGATGCA GAAAAGATC CCAAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAACTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAGCTG GAGACATGGA
1151 GACATTAACC TTCAGCTGAG AACTTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCC CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCCTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCTCTGT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTCAATCGT AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

BLAST Results

Entry HS188252 from database EMBL:

human STS WI-12265.

Score = 2554, P = 4.1e-109, identities = 556/587

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304
Category: putative protein


```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDSD
51  DDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GIDDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHMVKVSL EDAKNPKAID TWIESISELH RSKPPATVHY
201 TRMPDIDTL MQEWSPEFEE LLGKVSLEPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFALAE GKKAFTPSSN STSQAGDMET
301 LTFS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*)
Length = 562

HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:   3  DNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEEHGAPLEG 62
          +  DE EEE++ E++ T          +DDDDDDDD + D D DDD++E A G
Sbjct:  497 EEDDDDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

```

```

Query:   63 AYD 65
          D
Sbjct:  557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:   4  NSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEE 55
          N+ +E ++E+ +E + T + + N+DDDDDDDD + D D DDD++
Sbjct:  494 NNEEEDDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphut1_22n2, frame 3

Report for DKFZphut1_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION      1
[PROSITE]      CAMP_PHOSPHO_SITE  2
[PROSITE]      CK2_PHOSPHO_SITE  10
[PROSITE]      PKC_PHOSPHO_SITE   1
[PROSITE]      ASN_GLYCOSYLATION  3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY    11.84 %

```

```

SEQ  MADNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEEHGAPL
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  cccccchhhhhchhhhhcccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGIDDAFLKVP
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc

```

```

SEQ  RPDGKPDNLGLLVLEPSTKQSDPTVLSLWLTENSKQHNIQHMVKVSLDAEKNPKAID
SEG  .....
PRD  cccccccccceccccccccccccchhhhhccccccccccccccccchhhhhhhccccch

```

```

SEQ  TWIESISELHRSKPPATVHYTRMPDIDTLMQEWSPEFEEELGKVSLEPTAEIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhccccccccccccccccchhhhhccccchhhhhccccccccccccchhhhhh

```

```

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFALAEKGKAFTPSSNSTSQAGDMET
SEG  .....

```

Prosites for DKFZphute1_22n2.3

(No Pfam data available for DKFZphut1 22n2.3)

DKFZphut1_22o2

group: uterus derived

DKFZphut1_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="llp15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCGCGG GTCACAAGGA GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCTCAGGCC
151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCCGG CCCCCTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGCCAGGG CGGGGCCGCG GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCGGTCCCGC CTTCCTCCGC GCGCCATGGA GCCCGGGCG GTTGCAAGAG
351 CCGTGGAGAC GGGTGAGGAG GATGTGATTA TGGAAGCTCT GCGGTCATAC
401 AACCAGGAGC ACTCCAGAG CTTCACGTTT GATGATGCCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCCCA CCGTGTATC TGGGTGAGA GTGTCCGAAT CCTGTCCCGG
551 GACCGCAACT GCCTGGACCC GTTCACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCTGTCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCCC GCCTAGTGGT
751 GAACTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCACAG
801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGCTAAC GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCCTGTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGAACCCCC
951 CACCCACGCT CCTTCCTTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 AGGCTGAAGG GCGCTTTACC GACACCTGGG GACCTTCTC CGGCACCTGTG
1101 TGATGATCGC TACTGTGGA GACCGCACAG AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGAACTT GCCCTCAAG TGTCTGGATG TTCTCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGAGTG AATATGGATG
1251 TGATTCTGTC CCTCTCATC TTCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTCTG AGCGTGCTGA CTGAATGTGC
1351 CCGGATGCAC CGCCACGCA GGAAGTTCTT GAAGGCCAGG GGATGGCCAC
1401 CTCCCCAGGT GCTGCCCTCT CTGCGGGATG TGAGGACACG CCCTGAGGTT
1451 GGGGAGATGC TGCAGAACAA GCTTGTCCCG CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTCTCT GTTTGTCTCT TGCTCTGAGA
1551 GTGTGCCCGG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCCTAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCTGTG GACCGGGAGG GTGGAGGAGA AGCCGCCTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCAGCAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCGGGGGTCA TCTTACGTCC CTGCAAGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTTGTCTCC CCCATCAGGA CTGGTGTGTC TTCCAGAGAC TTCTTGGGG
2001 TTGCAACCTG GGAAGCCAC ATCCCACTGG ATCCACACCC GCGCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CGGTTCTGTT CATGATTTCG
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTTCA AAGTGTGTCG CTGAGGGCAG GGCCTACCTG
2251 GGCACACAGA AGAGCATATG GGAGGGCAGG GGTTTGGGTG TGGGTGCACA
2301 CAAAGCAGGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACTTCCCA GGGGAAAAAC CTGTGAGAAC TTTCATACG
2401 AGATATATCA AACACACCCT TCCAAGGTAT GTATGCTCTG TTGTTCTCTG
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCCTTGG
2501 TCCTCGTTCA GCTGCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAAACCTCC TTCTACTGCG TTTAGACTAC ACTTACAACA
2601 AGGAAATATG CCTCTGTGTG ACCATAGATT GAGATTTATA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```

BLAST Results

Entry AF015416 from database EMBL:
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:
human STS SHGC-15914.
Score = 1143, P = 9.0e-46, identities = 245/255

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEHQ QSFTFDDAQO EDRKRLAELL
51 VSVLEQGLPP SHRVIWLQSV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSPVESAD MDVVLESKLC LCNVLSSPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLFL LTALRTDVRQ QLFQELKGVV LLTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCVI IATAGDRTTE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRMLTHL DTDVKRVAE
401 FLFVLCSESV PRFIKYTYG NAAGLLAARG LMAGGRPEGQ YSEDEDTDT
451 EYKEAKASIN PVTGRVEEKP PNPMEGMTTE QKEHEAMKLV TMFDKLSRNR
501 VIQPMGMSPR GHLSLQDAM CETMEQQLSS DPDSDDPD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22o2, frame 2

TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7.
Length = 362

HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query: 215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRLRHCVMATAGDRTTEFHG 273
      SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct: 12 SQQNMVLTIELRLFPISKRSYLKEEDEQKILL-----LVIEIWASSLNNPNPSPLRW 65

Query: 274 HAVN-LLG-NLPLKCLDVLTLLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
      HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct: 66 HATNALLSFNLQLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query: 328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVTRTP-EVGEMLRNKLVR 386
      + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct: 122 QNTLPPILAILLLSLFFNIQNL-----SMLLFPTNDDRKQSLQKGKSFRCLLRL 173

Query: 387 MT-HLDTDKVRVAEFLFVLCSESVPRFIKYTYGNAAGLLAARGLMAGGRPEGQYS--- 442
      +T + + A L LC + + G G A G+ M P + +
Sbjct: 174 LTIPIVEPIGTYIASLLNELCDGDSQQIARIFGAGYAMGISQHSSETMPFSPLSKAASPV 233

Query: 443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMTTEEQKEHEAMKLVMTFDKLSRN 499
      + + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct: 234 FQKNSRGQENTEENLAIDPITGSMCTNRNKSQRLE-LSQEEKEREERLFYLFQRLEKN 292

```

Pedant information for DKFZphutel_22o2, frame 2

Report for DKFZphutel 22o2.2

```
[LENGTH] 537
[MW] 60372.53
[pI] 5.20
[BLOCKS] BL004151 Synapsins proteins
[PROSITE] MYRISTYL 4
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 9.50 %

SEQ MEPRVAEAVETGEEDVIMEALRSYNQEHQSQSFITDDAQQEDRKRLAELLVSVLEQGLPP
SEG .....
PRD ccchhhhhhhhhccchhhhhhhhhccccceecchhhhhhhhhhhhhhhhhhhccccc

SEQ SHRVIWLQSVRIILSRDNRCLDPFTSRQSLQALACYADISVSEGSVPESADMDVVLSESLKC
SEG .....
PRD ceeeeeeccccccccccccccccchhhhhhhhhhhhhcееееccccccccchhhhhhhhhhh

SEQ LCNLVLSSPVAQMLAAEARLVKLTERTVGLYRERSFPHDVQFFDLRLFLLLTALRTDVRQ
SEG .....
PRD hhhhhcccccchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhh

SEQ QLFQELKGVRLITDTLELTGLVTPEGNPPPTLLPSQETERAMEILKVLFNITLDSIKGEV
SEG .....
PRD hhhhhhchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccc

SEQ DEEDAALYRHLGTLRLHCVMIATAGDRTEEFHGHAVNLLGNLPLKCLDVLTLLEPHGDST
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccceeeeeeccccccc

SEQ EFMGVNMDVIRALLIFLEKRLHKTHRLKESVAPVLSVLTECARMHRPARKFLKAQGWWPP
SEG .....
PRD eeehhhhhhhhhhhhhhhhhhhhhhccccceeehhhhhhhhhhchhhhhhhhhhhccccccc

SEQ QVLPPLRDVTRPEVGEMLRNKLVRMLTHLTDVVKRAAEFLVLCSESVPFIKYTGYG
SEG .....
PRD cccccccccccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccceeeccc

SEQ NAAGLLAARGLMAGGRPEQGYSIEDTDTDEYKEAKASINPVTGRVEEKPPNPMEGMTEE
SEG .....
PRD xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
chhhhhhhhhccccccccccccccccccccchhhhhhhhhccccceeeccccccccchhhhh

SEQ QKEHEAMKLVTFMFKLSRNRVIQPMGMSPRGHLSLQDAMCETMEQQLSSDPDSDPD
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccc
```

Prosite for DKFZphut1 22o2.2

PS000001	230->234	ASN_GLYCOSYLATION	PDOC000001
PS000005	61->64	PKC_PHOSPHO_SITE	PDOC000005
PS000005	69->72	PKC_PHOSPHO_SITE	PDOC000005
PS000005	84->87	PKC_PHOSPHO_SITE	PDOC000005
PS000005	117->120	PKC_PHOSPHO_SITE	PDOC000005
PS000005	145->148	PKC_PHOSPHO_SITE	PDOC000005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000005	324->327	PKC_PHOSPHO_SITE	PDOC000005
PS000005	463->466	PKC_PHOSPHO_SITE	PDOC000005
PS000005	508->511	PKC_PHOSPHO_SITE	PDOC000005
PS000006	12->16	CK2_PHOSPHO_SITE	PDOC000006
PS000006	34->38	CK2_PHOSPHO_SITE	PDOC000006
PS000006	52->56	CK2_PHOSPHO_SITE	PDOC000006
PS000006	99->103	CK2_PHOSPHO_SITE	PDOC000006
PS000006	104->108	CK2_PHOSPHO_SITE	PDOC000006
PS000006	263->267	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006

WO 01/12659

PCT/IB00/01496

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphute1_22o2.2)

DKFZphute1_23e13

group: metabolism

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnological production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1 GGTATTATTA GCTCCTGGCT CCGCTCTAGA CCTCAGCGGT TCTGGCTGCC
51 AGCCTGGGCA GCCTGGGAAG CCTGGGAGGA CGGTGGCTTG CCGGTCTGTC
101 GTGAGGACGT GCGGACGGGG ACCCTCTGGG ATTCTGCTGG ATCTGCCCCG
151 GGGGTACCTT TTGGGGGCTG GGACCCAGT CGAGGGGACA CAACCGTCCC
201 TGGCAGTGGT TGGTTCTGCT TCTCCCTGCA GAAAAGCAGC ATTTTCGGAA
251 GCTGAAGAAAT AAGCTAGCCC AGCCACACCA CCTTGTGTG TGACCTTGGG
301 CAGGTGGTTC TGTCTCTCTG AGCCTCTGTT TCTCTCTGAG CTGAGCAGCC
351 ACCATGGCTG ACGGTGAGAT GCCCTTCTCC TGCCACTACC CAAGCCGCGT
401 GCGCCGAGAC CCTTCCGGG ACTTCCCTCT CTCTCTCGC CTGCTGGATG
451 ATGGCTTTGG CATGGACCCC TTCCAGACG ACTTGACAGC CTCTTGGCCC
501 GACTGGGCTC TGCTCTGCTC CTCTCCGCC TGGCCAGGCA CCCTAAGGTC
551 GGGCATGGTG CCCCGGGGCC CCACTGCCAC CGCCAGGTTT GGGGTGCCTG
601 CCGAGGGCAG GACCCCCCA CCTTCCCTG GGGAGCCCTG GAAAGTGTGT
651 GTGAATGTGC ACAGCTTCAA GCCAGAGGAG TTGATGGTGA AGACCAAAAG
701 TGGATACGTG GAGGTGTCTG GCAAACATGA AGAGAAACAG CAAGAAGGTG
751 GCATTGTTTC TAAGAACTTC ACAAAGAAAA TCCAGCTTCC TGCAAGGTTG
801 GATCCTGTGA CAGTATTTGC CTCACCTTCC CCAGAGGCTC TGCTGATCAT
851 CGAAGCTCCC CAGGTCCCTC CTTACTCAAC ATTGGAGAG AGCAGTTTCA
901 ACAACGAGCT TCCCCAGGAC AGCCAGGAAG TCACCTGTAC CTGAGATGCC
951 AGTACTGGCC CATCCTTGTG TTGTCCCCAA CCCTAGGGCT TCTCTGATTC
1001 CAGGATACAT TACTTTAGCT GAACCTCAGT TTAGTGCAAG TAAATGTGTA
1051 GAGGGTGCGG GGGTGAGGAC TGACCACAGA TTCCCTGGAT AGTGTAGTGG
1101 TAGATTCTC CACAGGATAG CGCAATTGGC AAATCATGCT TGGTTGTGTT
1151 AGGCCAAAAT ACTAGTTTTC CTTTCTTAC CTTTCTATC TTGATGAAAA
1201 TGTGTGCAT TCTATAGTTG CAAAACACAT AAAAGGGGAC TTAACATTTT
1251 ACGTTGTATC TTACTTGAGC TGAATGCAAG GGTACTTTT CTCTGGGGAC
1301 CTCCCCATC ACCCAGGTTT CTACTCTGGG CTCCCGATTG CCATGGCTCC
1351 CAAACCATGC CGCATGGTTT GGTAAATGAA ACCCAGTAGC TAACCCCACT
1401 GTGCTTCCAC ATGCTCGGCC TAAATGGGT GATATACAGG TCTTATATCC
1451 CCATATGGAA TTTATCCATC AACCACATAA AAACAAACAG TGCTTCTGTC
1501 CCTCTGCCCA GATGTGTCCA GCACGTTCTC AAAGTTTCCA CATTAGCACT
1551 CCCTAAGGAC GCTGGGAGCC TGTCACTTTA TGATCTGACC TAGGTCCCCC
1601 CTTTCTTCTG TCCCCTGTGT TTAAGTCGGG ATTTTACAG AGGGAGCTGT
1651 CTCCAGACAG CTCATCAGG AACCAAGCAA AGGCCAGATA GCCTGACAGA
1701 TAGGCTAGTG GTATTGTGTA TATGGGCGGG ACGTGTGTGT CATTATTATT
1751 TGAGTTATGC TGTGTGTTAG GGGTAAATAA CAGTAAATAA TTAATAATAA
1801 TAATAATAAT AATAAAGGAG CTGACGTTCT TAAAAAGAA AAAAAAATAA
1851 AAAA
```

BLAST Results

Entry HS286348 from database EMBL:
human STS TIGR-A002J47.
Score = 510, P = 1.2e-16, identities = 102/102

Medline entries

95394379:
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:
Physiological and pathological changes in levels of the two
small stress proteins, HSP27 and alpha B crystallin, in rat
hindlimb muscles

Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196
Category: strong similarity to known protein
Prosites motifs: SUBTILASE_ASP (28-39)

1 MADGQMPFSC HYPSRLRRDP FRDSPLSSRL LDDGFGMDPF PDDLTASWPD
51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGEPWKVCV
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQO EGGIVSKNFT KKIQLPAEVD
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P =
4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561_1 product: "heat shock protein HSP27"; Mus musculus
heat shock protein HSP27 internal deletion variant b mRNA, complete
cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog
Length = 209

HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27
Identities = 80/182 (43%), Positives = 102/182 (56%)

Query: 1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRLDDGFGMDPFDDLTASWPDWALPRLSS 58
M + ++PFS PS DPFRD P SRL D FG+ P++ W W S
Sbjct: 1 MTERRVPFSLRSPSW---DPFRDWYPAHSRLFDQAFGLPRLPEE---WAQWFG---HS 50

Query: 59 AWPGLRSGMVP---RGPTATARFGVPAEGR--TPPPFPG-----EPWKVCVNVHSF 105
WPG +R +P GP A A PA R + G + W+V ++V+ F
Sbjct: 51 GWPGYVRP--IPPAVEGPAAAAAAPAYSRLSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query: 106 KPEELMVKTKDGYVEVSGKHEEKQEGGIVSKNFTKKIQLPAEVDVPTVFASLSPEGLLI 165
PEEL VKTKDG VE++GKHEE+Q E G +S+ T K LP VDP V +SLSPEG L
Sbjct: 109 APEELTVKTKDGVVEITGKHEERQDEHGYISRRLTPKYTLPPGVDPTLVSSSLSPGTLT 168

Query: 166 IEAPQVPPYSTFGE 179
+EAP P + E
Sbjct: 169 VEAPMPKPKATQSAE 182

Pedant information for DKFZphut1_23e13, frame 3

Report for DKFZphut1_23e13.3

[LENGTH] 196
[MW] 21604.37


```

[pI]          5.00
[HOMOL]       PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS]      BL01031C
[PIRKW]       blocked amino end 1e-13
[PIRKW]       acetylated amino end 4e-13
[PIRKW]       phosphoprotein 7e-21
[PIRKW]       glycoprotein 2e-11
[PIRKW]       heat shock 7e-21
[PIRKW]       molecular chaperone 4e-13
[PIRKW]       alternative splicing 1e-19
[PIRKW]       eye lens 6e-14
[PIRKW]       stress-induced protein 7e-21
[SUPFAM]      alpha-crystallin 7e-21
[PROSITE]     SUBTILASE_ASP 1
[PROSITE]     MYRISTYL 2
[PROSITE]     CK2_PHOSPHO_SITE 2
[PROSITE]     PKC_PHOSPHO_SITE 6
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        Heat shock hsp20 proteins
[KW]          All_Beta
[KW]          LOW_COMPLEXITY 7.14 %

```

```

SEQ  MADGQMFFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTA SWPDWALPRLSSAW
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PGTLRSGMVPFRGPTATARFGVPAEGRTPPPFPGE PWKVCNVVHSFKPEELMVKT KDGYVE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VSGKHEEKQEGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLIIEAPQVPPYSTFGES
SEG  .....
PRD  eccchhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFNNELPQDSQEV TCT
SEG  .....
PRD  ccccccccccccccccc

```

Prosite for DKFZphute1_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

Pfam for DKFZphute1_23e13.3

```

HMM_NAME      Heat shock hsp20 proteins
HMM            *AMMrpPWDWRE....DpDHFeVrMDMPGFKPEEIKVkvEDNNVLvIeG
               A   P++ R   + ++V++++ FKPEE+ VK+ D+ +++++G
Query         77  ARFGVPAEGR-TPPPFPGE PWKVCNVVHSFKPEELMVKT KDG-YVEVSG 123

HMM            EHEREEEREDDKWWHERIYRHFMRFRrLPENVDPDqIkAsMSdNGVLTl
               +HE E++   + + ++ F ++LP +VDP + AS+S++G+L I
Query         124 KHE---EKQQ----EGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLI 166

HMM            TVPKpEP*
               ++P ++P
Query         167 EAPQVPP 173

```

DKF2phutel_23g11

group: uterus derived

DKF2phutel_23g11 encodes a novel 256 amino acid protein with similarity to S.pombe SPAC31G5.12c and S. cerevisiae Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GCGGTCGCGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGGATGGG CCGGCAGGGC
151 CGGGAGTAAC GGGACGTCGC CGCGGAGCTT CTCCCCCGG ATACAGTGCG
201 GCCCGAGCGG AGGCCGCGGC GCCGCCCTCC GATCTTGAAG AGCCCCGCGT
251 GCGCGGAGCC CGCCCCCGCC TCGCGACCGG CACCGACGCG GAGCGACCCG
301 CCCAGCCAGA CCCGGCCCGG CGCGGCTGA TCTAACCCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAG ACATGAAGCT
401 ATTGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGTACTC ATGTAAGATG
501 GCAGGAGACG ACAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACCTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAAGCCAA GCGCGTGAGG AGGAGGGCCC CCTCAGTGAC
651 AAGTGCAGCC GCAAGACCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTGAGCC
751 GGGAGGCCAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCCTC
951 AACTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGTTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGCG CTCCCCAGCT
1251 GCTGGCCAGA CCTTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCTAGC CTTTGGCTC CATCTGTGG ATGCCCACTC ACCCTCAGA
1351 CTCTGCTGC CCAATGCTGTG GCCGGACTTG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGGAGCT GCCCTGCCCA AATGAACTGC CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCCAAA
1501 GGTCTGTGG CCGGAGGCC CACGAGCAG CCCAGCAGT CACCGGCTCT
1551 GGTCTGGGC GGGCCCCGGT GCCCACCTGT ACCCCACCT CGCCCATTTG
1601 GCCGCTGCA CTGAGTGCA CTTTGCTGCA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256
Category: similarity to known protein

```

1 MKLLENSFE AINSQTLVET GDAHIIGRIE SYSCKMAGDD KHMFKQFCQE
51 GQPHVLEALS PPQTSGLSPS RLSKSQGGEE EGPLSDKCSR KTLFYLIATL
101 NESFRPDYDF STARSHEFSR EPSLSWVVNA VNCSLFSAVR EDFKDLKPQL
151 WNAVDEEICL AECDIYSYNP DLDSDPFGED GSWLSFNYFF YNKRLKRIVF
201 FSCRSISGST YTPSEAGNEL DMELGEEVEE EESRSRGSGA EETSTMEEDR
251 VPVICI

```

BLASTP hits

Entry SPAC31G5_12 from database TREMBL:
 gene: "SPAC31G5.12c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c31G5.
 Score = 272, P = 9.3e-24, identities = 51/127, positives = 80/127

Entry SPD656_1 from database TREMBL:
 product: "ORF N150"; Yeast DNA for bfr2+ protein/pad1+ protein/sks1+
 protein, ORF N313, ORF N150, complete cds, and for ORF N118, partial
 cds.
 Score = 263, P = 8.4e-23, identities = 50/127, positives = 79/127

Entry S50986 from database PIR:
 MAF1 protein - yeast (Saccharomyces cerevisiae) >SWISSPROT:MAF1_YEAST
 MAF1 PROTEIN. >TREMBL:SC19492_1 gene: "MAF1"; product: "Maf1p";
 Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds.
 >TREMBL:SC8119_11 gene: "MAF1p"; product: "Maf1p"; S.cerevisiae
 chromosome IV cosmid 8119.
 Score = 180, P = 2.3e-17, identities = 43/133, positives = 75/133

Entry AF098499.2 from database TREMBL:
 gene: "C43H8.2"; Caenorhabditis elegans cosmid C43H8.
 Score = 263, P = 9.2e-23, identities = 78/252, positives = 118/252

Alert BLASTP hits for DKFZphut1_23g11, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_23g11, frame 3

Report for DKFZphut1_23g11.3

```

[LENGTH]      256
[MW]           28869.95
[pI]           4.51
[HOMOL]        TREMBL:SPAC31G5_12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR005c]
6e-13
[PROSITE]      MYRISTYL          3
[PROSITE]      CK2_PHOSPHO_SITE    5
[PROSITE]      PKC_PHOSPHO_SITE    6
[PROSITE]      ASN_GLYCOSYLATION   3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      7.81 %

```

```

SEQ  MKLLENSFEAINSQLTVETGDAHIIGRIESYSCKMAGDDKHMFKQFCQEGQPHVLEALS
SEG  .....
PRD  cccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhcccccccccccc

```

```

SEQ  PPQTSGLSPSRLSKSQGGEEEGPLSDKSRKTLFYLIATLNESEFRPDYDFSTARSHEFSR
SEG  .....
PRD  cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhcccccccccccccccc

```

```

SEQ  EPSLSWVVNAVNCSLFSAVREDFKDLKPQLWNAVDEEICLAECDIYSYNPDLDSDPFGED
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc

```

```

SEQ  GSWLSFNYFFYNKRLKRIVFFSCRSISGSTYTPSEAGNELDMELGEEVEEESRSRGSGA
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  EETSTMEEDRVPVICI
SEG  xx.....
PRD  ccccccccccccccccc

```

Prosite for DKFZphut1_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_23g11.3)

DKFZphut1_24c19

group: transmembrane protein

DKFZphut1_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown

membrane regions: 1

Summary DKFZphut1_24c19 encodes a novel 195 amino acid protein, with no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp

Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
1 ACGAGTCAGC CAAAGATGGC TGCGCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACCTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAACAGT CTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTTCT ATTTTGCTCC AGACTATGTT TTCAGCATAC CTGGGTCTG
651 AACAATATAA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCACG GATTTTAAAC AAATATGTAA ACAAATAAAT AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195

Category: putative protein

```
1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51 NSLFRRILNV TKARIAAGLP MAGIPFLTDD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGEQYKL LIKALQLSEP GKEIH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1_24c19, frame 2

Report for DKFZphut1_24c19.2

```

[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSLFRILNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPMAGIPFLTTLTYRCFVSFPLNTGDLDCETCTITRSGTLGLVIGGLYPVF
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPIILLQTMFSAYLGSEQYKL
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....

```

Prosite for DKFZphut1_24c19.2

```

PS00001      11->15  ASN_GLYCOSYLATION      PDOC00001
PS00001      34->38  ASN_GLYCOSYLATION      PDOC00001
PS00001      59->63  ASN_GLYCOSYLATION      PDOC00001
PS00005      18->21  PKC_PHOSPHO_SITE      PDOC00005
PS00005      82->85  PKC_PHOSPHO_SITE      PDOC00005
PS00005     151->154 PKC_PHOSPHO_SITE      PDOC00005
PS00006      13->17  CK2_PHOSPHO_SITE      PDOC00006
PS00008      40->46  MYRISTYL      PDOC00008
PS00008      47->53  MYRISTYL      PDOC00008
PS00008      68->74  MYRISTYL      PDOC00008
PS00008     110->116 MYRISTYL      PDOC00008
PS00008     127->133 MYRISTYL      PDOC00008
PS00008     142->148 MYRISTYL      PDOC00008

```

(No Pfam data available for DKFZphut1_24c19.2)

DKFZphute1_24e11

group: intracellular transport and trafficking

DKFZphute1_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits

potential start at 184,

TRANSMEMBRANE 4

function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp

Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
1  ACGCGTCCGG CAGAAGCTCG GAGCTCTCGG GGTATCGAGG AGGCAGGCC
51  GCGGGCGCAC GGGCGAGCGG GCCGGGAGCC GGAGCGGCGG AGGAGCGCGG
101 AGCAGCGCGC CGGCGGGCTC CAGGCGAGGC GGTGACGCTC CTGAAAAC
151 TGCGCGCGCG CTCGCGCCAC TCGCGCCGGA GCGATGAAGA TGGTCGCGCC
201 CTGGACGCGG TTCTACTCCA ACAGCTGCTG CTTGTGCTGC CATGTCCGCA
251 CCGGCACCAT CTGCTCGGCG GTCTGGTATC TGATCATCAA TGCTGTGGTA
301 CTGTTGATTT TATTGAGTGC CCTGGCTGAT CCGGATCAGT ATAACTTTTC
351 AAGTTCTGAA CTGGGAGGTG ACTTTGAGTT CATGGATGAT GCCAACATGT
401 GCATTGCCAT TGGGATTTCT CTTCTCATGA TCCTGATATG TGCTATGGCT
451 ACTTACGGAG CGTACAAGCA ACGCGCAGCC TGGATCATCC CATCTTCTCG
501 TTACACAGAT TTTGACTTTG CCTGAACAT GTTGGTTGCA ATCACTGTGC
551 TTATTTATCC AAATCCATT CAGGAATACA TACGGCAACT GCCTCCTAAT
601 TTTCCCTACA GAGATGATGT CATGTCAGTG AATCCTACCT GTTTGGTCCT
651 TATTATTCTT CTGTTTATTA GCATTATCTT GACTTTTAAG GGTACTTGA
701 TTAGCTGTGT TTGGAATGCG TACCGATACA TCAATGGTAG GAACTCCTCT
751 GATGTCCTGG TTTATGTTAC CAGCAATGAC ACTACGGTGC TGCTACCCCC
801 GTATGATGAT GCCACTGTGA ATGGTGCTCG CAAGGAGCCA CCGCCACCTT
851 ACGTGTCTGC CTAAGCCTTC AAGTGGGCGG AGCTGAGGGC AGCAGCTTGA
901 CTTTGCAGAC ATCTGAGCAA TAGTTCTGTT ATTTCACTTT TGCCATGAGC
951 CTCTCTGAGC TTGTTTGTG CTGAATGCT ACTTTTAAA ATTTAGATGT
1001 TAGATTGAAA ACTGTAGTTT TCAACATATG CTTTGCTAGA ACACGTGTGAT
1051 AGATTAACTG TAGAATCTTT CCTGTACGAT TGGGGATATA ACGGGCTTCA
1101 CTAACCTTCC CTAGGCATTG AAATCTCCC CAAATCTGAT GGACCTAGAA
1151 GTCTGCTTTT GTACCTGCTG GGGCCCAAAG TTGGGCATTT TTCTCTCTGT
1201 TCCCTCTCTT TTGAAAATGT AAAATAAAAC CAAAAATAGA CAACTTTTTC
1251 TTCAGCCATT CCAGCATAGA GAACAAACC TTATGGAAAC AGGAATGTCA
1301 ATTTGTGAAT CATTGTTCTA ATTAGGTAAA TAGAAGTCCT TATGTATGTG
1351 TTACAAGAAAT TTCCCCACAC ACATCCTTTA TGAAGTGAAT TCAATGACAG
1401 TTTGTGTTTG GTGTAAGG ATTTCTCCA TGGCTGAAT TAAGACCATT
1451 AGAAAGCACC AGGCGTGGG AGCAGTGACC ATCTACTGAC TGTTCTTGTG
1501 GATCTTGTGT CCAGGGACAT GGGGTGACAT GCCTCGTATG TGTAGAGGG
1551 TGGAAATGGAT GTGTTTGGCG CTGCATGGGA TCTGGTGCCC CTCTTCTCCT
1601 GGATTACATC CCCACCCAG GGGCCGCTTT TACTAAGTGT TCTGCCCTAG
1651 ATTTGTTCAA GGAGGTCATC CAACTGACTT TATCAAGTGG AATTGGGATA
1701 TATTTGATAT ACTTCTGCTT AACAACTGG AAAAGGGTTT TCTTTTCCCT
1751 GCAAGCTACA TCCTACTGCT TTGAACCTCC AAGTATGTCT AGTCACCTTT
1801 TAAAATGTAA ACATTTTCAG AAAAATGAGG ATTGCCTTCC TTGTATGCGC
1851 TTTTACCTT GACTACCTGA ATTGCAAGG ATTTTATAT ATTCATATGT
1901 TACAAAGTCA GCAACTCTCC TGTGGTTCA TTATTGAATG TGCTGTAAAT
1951 TAAGTCGTTT GCAATTAAAA CAAGGTTTGC CCACATCCAA AAAAAAAAAA
2001 AAAAA
```

BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.
Score = 1629, P = 3.1e-67, identities = 343/354

Medline entries

96199248:
Identification of a novel membrane transporter
associated with intracellular membranes by
phenotypic complementation in the yeast
Saccharomyces cerevisiae.

Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226
Category: strong similarity to known protein

1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP
51 DQYNFSSSEL GGDFFEMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW
101 IIPFFCYQIF DFLNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN
151 PTCLVLIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT
201 TVLPPPYDDA TVNGAAKEPP PPYVSA

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24e11, frame 1

SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981_1 product: "E3 protein"; Human retinoic acid-inducible
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108).
Length = 233

HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53
Identities = 102/221 (46%), Positives = 148/221 (66%)

Query: 9 RFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSELGGDF- 64
RFYS CC CCHVRTGTI+LG WY+++N ++ ++L + P+ N +G +
Sbjct: 13 RFYSTRCCGCCHVRTGTIILGTWYVMVNNLLMAILLTVEVTHPNSMPAVNIQYEVIGNYYS 72
Query: 65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123
E M D N C+ A+S+LM +I +M YGA + W+IPFFCY++FDF L+ LVAT+ L
Sbjct: 73 SERMAD-NACVLFVAVSVLMFISSMLVYGAISYQVQWLIIPFFCYRLDFVLSCLVAISSL 131
Query: 124 IYPNSIQEYIRQLPPNFYRDDVMSVNPCLVLIILLFISIILTFKGYLISCVWNCYRYI 183
Y I+EY+ QLP +FPY+DD+++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI
Sbjct: 132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCWNCYKYI 190
Query: 184 NGRNSSDVLVYVTSN-DTTVLLPPYDDATVNGAAKEPPPPYVSA 226
N RN ++ VY +LP Y+ A V KEPPPPY+ A
Sbjct: 191 NNRNVPEIAVYPAFEAPPQYVLPTYEMA-VKMPEKEPPPPYLPA 233

Pedant information for DKFZphut1_24e11, frame 1

Report for DKFZphut1_24e11.1

[LENGTH] 226
[MW] 25419.11

[pI] 4.65
 [HOMOL] SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).
 5e-40
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 1
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] SIGNAL PEPTIDE 49
 [KW] TRANSMEMBRANE 2
 [KW] LOW_COMPLEXITY 20.80 %

SEQ MKMVAPWTRFYSNSCLCCHVRTGTILLGVWYLLINAVVLLILLSALADPDQYNFSSSEL
 SEGXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEM

SEQ GGDFEFMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFCYQIFDFALNMLVAI
 SEGXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEMMM

SEQ TVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIIILTFKGYLISCVWNCY
 SEGXXXXXXXXXXXXXXXXX.....
 PRD hhhcc
 MEM MMMMM.....MM...

SEQ RYINGRNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA
 SEG
 PRD ecc
 MEM

Prosites for DKFZphut1_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphut1_24e11.1)

DKF2phutel_24j6

group: cell structure and motility

DKF2phutesl_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CARL).

The novel protein is very similar to Carl and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CARL A.thaliana T19C21.5

complete cDNA, complete cds, EST hits
potential frame shift at Bp 1241 according to CARL
but frame shift might be in CARL sequence!
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```
1  ACGCGTCCGA  GCTGGCTCAG  GCGCTCCGCT  AGGCTCGGAC  GACCTGCTGA
51  GCCTCCCAAA  CCGCTTCCAT  AAGGCTTTGC  CTTTCCAACT  TCAGCTACAG
101  TGTTAGCTAA  GTTTGGAAAG  AAGGAAAAAA  GAAAAATCCCT  GGGCCCCCTT
151  TCTTTTGTTC  TTTGCCAAAG  TCGTCGTTGT  AGTCTTTTGT  CCCAAGGCTG
201  TTGTGTTTTT  AGAGGTGCTA  TCTCCAGTTC  CTGCACTCC  TGTAAACAAG
251  CACCTCAGCG  AGAGCAGCAG  CAGCGATAGC  AGCCGAGAA  GAGCCAGCGG
301  GGTCCGCTAG  TGTATGACC  AGGCGGGAG  ATCACAACCG  CCAGAGAGGA
351  TGCTGTGGAT  CCTTGGCCGA  CTACCTGACC  TCTGCAAAAT  TCCTTCTCTA
401  CCTTGGTCAT  TCTCTCTCTA  CTGGGGGAGA  TCGGATGTGG  CACTTTGCGG
451  TGTCGTGTTT  TCTGGTAGAG  CTCTATGAA  ACAGCCTCCT  TTTGACAGCA
501  GTCTACGGGC  TGGTGGTGCC  AGGGTCTGTT  CTGGTCTGG  GAGCCATCAT
551  CGGTGACTGG  GTGGACAAGA  ATGCTAGACT  TAAAGTGGCC  CAGACCTCGC
601  TGGTGGTACA  GAATGTTTCA  GTCATCCTGT  GTGGAATCAT  CCTGATGATG
651  GTTTTCTTAC  ATAAACATGA  GCTTCTGACC  ATGTACCATG  GATGGGTCTT
701  CACTTCTTGC  TATATCCTGA  TCATCACTAT  TGCAAAATAT  GCAAAATTGG
751  CCAGTACTGC  TACTGCAATC  ACAATCCAAA  GGGATTGGAT  TGTGTTGTTT
801  CAGAGAGAAG  ACAGAAGCAA  ACTAGCAAA  ATGAATGCCA  CAATACGAAG
851  GATTGACGAG  TTAACCAACA  TCTTAGCCCC  CATGGCTGTT  GGCCAGATTA
901  TGACATTGGG  CTCCCCAGTC  ATCGGCTGTG  GCTTTATTTC  GGGATGGAAC
951  TTGGTATCCA  TGTGCGTGGA  GTACGTCTGT  CTCTGGAAGG  TTTACAGAA
1001  AACCCAGCT  CTAGCTGTGA  AAGCTGGTCT  TAAAGAAGAG  GAAACTGAAT
1051  TGAACAGCT  GAATTTACAC  AAAGATACTG  AGCCAAAACC  CCTGGAGGGA
1101  ACTCATCTAA  TGGGTGTGAA  AGACTCTAAC  ATCCATGAGC  TTGAACATGA
1151  GCAAGAGCCT  ACTTGTGCCT  CCCAGATGGC  TGAGCCCTTC  CGTACCTTCC
1201  GAGATGGATG  GGTCTCCTAC  TACAACAGC  CTGTGTTTCT  GGCTGGCATG
1251  GGTCTTGCTT  TCCTTTATAT  GACTGTCTGT  GGCTTTGACT  GCATCACCAC
1301  AGGGTACGCC  TACATCAGG  GACTGAGTGG  TTCCATCCTC  AGTATTTTGA
1351  TGGGAGCATC  AGCTATAACT  GGAATAATGG  GAAGTGTAGC  TTTTACTTGG
1401  CTACGTCGAA  AATGTGGTGT  GGTCCGACA  GGTCTGATCT  CAGGATTGGC
1451  ACAGCTTTCC  TGTGTGATCT  TGTGTGTGAT  CTCTGTATTC  ATGCCTGGAA
1501  GCCCCTTGA  CTGTCCGTT  TCTCCTTTG  AAGATATCCG  ATCAAGGTTT
1551  ATCAAGGAG  AGTCAATTAC  ACCTACCAAG  ATACCTGAAA  TTACAACATG
1601  AATATACATG  TCTAATGGGT  CTAATTCTGC  TAATATTGTC  CCGGAGACAA
1651  GTCCTGAATC  TGTGCCATA  ATCTCTGTCA  GTCTGCTGTT  TGCAGGCGTC
1701  ATTGCTGCTA  GAATCGGTCT  TTGGTCCTTT  GATTTAACTG  TGACACAGTT
1751  GCTGCAAGAA  AATGTAATTG  AATCTGAAAG  AGGCATTATA  AATGGTGTAC
1801  AGAAGTCCAT  GAACTATCTT  CTGTATCTTC  TGCATTTCAT  CATGGTCATC
1851  CTGGCTCCAA  ATCCTGAAGC  TTTTGGCTTG  CTGATTTGA  TTTCACTCTC
1901  CTTTGTGGCA  ATGGGCCACA  TTATGTATTT  CCGATTTCGC  CAAAATACTC
1951  TGGGAAACAA  GCTCTTTGCT  TGCGGTCTGT  ATGCAAAAGA  AGTTAGGAAG
2001  GAAATCAAG  CAAATACATC  TGTGTTTGA  GACAGTTTAA  CTGTTGCTAT
2051  CCTGTACTA  GATTATATAG  AGCAGATGTG  CTTATTTTGT  ACTGCAGAA
2101  TCCAATAAAT  GGCTGGGTGT  TTTGCTCTGT  TTTTACCACA  GCTGTGCCCT
2151  GAGAACTAAA  AGCTGTTTAG  GAAACCTAAG  TCAGCAGAAA  TAACTGATT
2201  AATTTCCCTT  ATGTTGAGGC  ATGGAAAAAA  AATTGGAAAA  GAAAACTCA
2251  GTTTAAATAC  GGAGACTATA  ATGATAACAC  TGAATTCCCC  TATTTCTCAT
2301  GAGTAGATAC  AATCTTACGT  AAAAGAGTGG  TTAGTCACGT  GAATTCAGTT
2351  ATCATTTGAC  AGATTCTTAT  CTGTACTAGA  ATTCAGATAT  GTCAGTTTTC
2401  TGCAAAACTC  ACTCTTGTTC  AAGACTAGCT  AATTTATTTT  TTTGCATCTT
2451  AGTTATTTT  AAAACAAAT  TCTTCAAGTA  TGAAGACTAA  ATTTTGATA
2501  CTAATATTAT  CCTATTGAT  CCTATTGATC  TTAAGGTATT  TACATGTATG
```

```

2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAAGCACT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAG CTGTAAAGTA GATTATACAC ACAATGACTG CATACAGACT
2751 TCAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACCTCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCCACA CTTGTGAAGG TTTTGTTTA CAAATCACTT GATTAAACAC
2951 ACTCAGGTAG AATATTTTAA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAGTC AGTTTGCAAC ATGCTGTGAC CAAGATGGTA CTTGCGCTTA
3151 ACCGTTTATA TGCACCTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TCTTTTATCC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA

```

BLAST Results

Entry HS389210 from database EMBL:
human STS SHGC-10164.
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:
human STS WI-16551.
Score = 1193, P = 5.7e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571
Category: strong similarity to known protein

```

1 MTRAGDHNRO RGCCGSLADY LTSAKFLLYL GHSLSTWGRD MWHFAVSVFL
51 VELYGNLLLL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKHEL LTMVHGWLVT SCYILITIA NIANLASTAT
151 AITIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNIALPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCWEY VLLWKVYQRT PALAVKAGLK EEETELQQLN
251 LHKDTEPKPL EGTLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDQWV
301 SYYNQPVFLA GMGLAFLYMT VLGFDICITG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAF TWLRRRCGLV RTGLISGLAQ LSCLILCVIS VFMPGSPDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEI YMSNGSNSAN IVPETSPESV
451 PIISVSLLEA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLVLVLSVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_24j6, frame 3

TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P
= 2.8e-60

TREMBL:AF039046_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.
Length = 405

HSPs:

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151
Identities = 288/319 (90%), Positives = 297/319 (93%)

```
Query: 1 MTRAGDHNRRQCCGSLADYLSAKFLLYLGHSLSTWGDMMHFAVSVFLVELYGNLSLL 60
      MT++ D Q GCCGSLA+YLTSKFLLYLGHSLSTWGDMMHFAVSVFLVELYGNLSLL
Sbjct: 1 MTKSRDQTHQEGCCGSLANYLSAKFLLYLGHSLSTWGDMMHFAVSVFLVELYGNLSLL 60

Query: 61 TAVYGLVVAGSVLVLGAIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL 120
      TAVYGLVVAGSVLVLGAIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHK+EL
Sbjct: 61 TAVYGLVVAGSVLVLGAIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKNEL 120

Query: 121 LTMVHGWVLTSCYILITIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI 180
      L MYHGWVLT CYILITIANIANLASTATAITIQRDWIVVAGE+RS+LA+MNATIRRI
Sbjct: 121 LNMVHGWVLTVCYILITIANIANLASTATAITIQRDWIVVAGENRSRLADMNATIRRI 180

Query: 181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK 240
      DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEY LLWKVYQKTPALAVKA LK
Sbjct: 181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYFLLWKVYQKTPALAVKAALK 240

Query: 241 EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVW 300
      EE+ELKQL KDTEPKPLEGTHLMG KDSNI ELE EQEPTCASQ+AEFPRTFRDGVW
Sbjct: 241 VEESELKQLTSKDPKTEPKPLEGTHLMGEKDSNIRELECEQEPTCASQIAEPFRTFRDGVW 300

Query: 301 SYYNQPVFLAGMGLAF-LY 318
      SYYNQPVFL G F LY
Sbjct: 301 SYYNQPVFLGWHGPGFPLY 319
```

Pedant information for DKFZphute1_24j6, frame 3

Report for DKFZphute1_24j6.3

```
[LENGTH] 571
[MM] 62542.72
[PI] 6.08
[HOMOL] TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator (CAR1) mRNA, complete cds. 1e-141
[BLOCKS] BL00341D
[PROSITE] MYRISTYL 15
[PROSITE] MITOCH CARRIER 1
[PROSITE] CK2 PHOSPHO SITE 6
[PROSITE] PROKAR LIPOPROTEIN 1
[PROSITE] PKC PHOSPHO SITE 4
[PROSITE] ASN GLYCOSYLATION 4
[PFAM] Laminin B (Domain IV)
[KW] TRANSMEMBRANE 4
[KW] LOW_COMPLEXITY 8.76 %
```

```
SEQ MTRAGDHNRRQCCGSLADYLSAKFLLYLGHSLSTWGDMMHFAVSVFLVELYGNLSLL
SEG .....
PRD cccccccccccccchhhhhhhheeeccceccccchhhhhhhheeecccccce
MEM .....MMMMMMMMMMMM

SEQ TAVYGLVVAGSVLVLGAIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG .xxxxxxxxxxxxxxxxxxxxx
PRD ehhhhhhhhccceeeccccchhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhh
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ LTMVHGWVLTSCYILITIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI
SEG .....xxxxxxxxxxxxxxxxxxxxx
PRD hhccccchhhhhhhhhhhhhhhhhhhhhheeeccceeeccccchhhhhhhhhhh
MEM MMMMM.....

SEQ DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG .....
PRD hhhhhhccceeeceeeceeeceeeccchhhhhhhhhhhhhccchhhhhhhhh
MEM .....

SEQ EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVW
SEG .....
PRD hhhhhhhhhccceccccccccceeecccccceccccccccccccccccccccce
MEM .....

SEQ SYYNQPVFLAGMGLAFLYMTVLGFDCTTGAYTQGLSGSILSILMGASAITGIMGTVAF
SEG .....
PRD eeccccceccccchhhhhhhccccceeececcccceeeccccceeeccccceehhhhhh
```

```

MEM      .....

SEQ      TWLRKCKGLVRTGLISGLAQLSCILCVISVFMGPSPLDLSVSPFEDIRSRFIQGESITP
SEG      xxxxxxxxxx
PRD      hhhhhhhccccccccchhhhhhhhhhhhhhhhhccccccccccccchhhhhhhcccccccc
MEM      .....

SEQ      TKIPEITTEIYMSNGSNSANIVPETSPEVPIISVSLLFAGVIAARIGLWSFDLTVTQLL
SEG      xxxxxxxxxx
PRD      cccccceeeeeeccccccccccccccccccccceeeeee hhhhhhhhhhhccccchhhhhhhhh
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      QENVIESERGIINGVQNSMYNLDDLHFIMVILAPNEAFGLLVLSVSFVAMGHIMYFR
SEG      .....
PRD      hhhhhccccceeeeeeccccchhhhhhhhhhhheeeccccccccceeeeeeccccccccceee
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      FAQNTLGNKLFACGPDACEVVRKENQANTSVV
SEG      .....
PRD      eccccccccceeeccccchhhhhhhhhcccccc
MEM      .....

```

Prosite for DKFZphute1 24j6.3

PS000001	100->104	ASN_GLYCOSYLATION	PD0C000001
PS000001	174->178	ASN_GLYCOSYLATION	PD0C000001
PS000001	434->438	ASN_GLYCOSYLATION	PD0C000001
PS000001	567->571	ASN_GLYCOSYLATION	PD0C000001
PS000005	23->26	PKC_PHOSPHO_SITE	PD0C000005
PS000005	176->179	PKC_PHOSPHO_SITE	PD0C000005
PS000005	294->297	PKC_PHOSPHO_SITE	PD0C000005
PS000005	487->490	PKC_PHOSPHO_SITE	PD0C000005
PS000006	16->20	CK2_PHOSPHO_SITE	PD0C000006
PS000006	36->40	CK2_PHOSPHO_SITE	PD0C000006
PS000006	294->298	CK2_PHOSPHO_SITE	PD0C000006
PS000006	396->400	CK2_PHOSPHO_SITE	PD0C000006
PS000006	403->407	CK2_PHOSPHO_SITE	PD0C000006
PS000006	445->449	CK2_PHOSPHO_SITE	PD0C000006
PS000008	12->18	MYRISTYL	PD0C000008
PS000008	65->71	MYRISTYL	PD0C000008
PS000008	76->82	MYRISTYL	PD0C000008
PS000008	193->199	MYRISTYL	PD0C000008
PS000008	267->273	MYRISTYL	PD0C000008
PS000008	311->317	MYRISTYL	PD0C000008
PS000008	336->342	MYRISTYL	PD0C000008
PS000008	339->345	MYRISTYL	PD0C000008
PS000008	353->359	MYRISTYL	PD0C000008
PS000008	368->374	MYRISTYL	PD0C000008
PS000008	373->379	MYRISTYL	PD0C000008
PS000008	435->441	MYRISTYL	PD0C000008
PS000008	461->467	MYRISTYL	PD0C000008
PS000008	490->496	MYRISTYL	PD0C000008
PS000008	494->500	MYRISTYL	PD0C000008
PS000103	122->133	PROKAR_LIPOPROTEIN	PD0C000103
PS002115	404->414	MITOCH_CARRIER	PD0C001899

Pfam for DKFZphute1 24j6.3

HMM_NAME	Laminin B (Domain IV)		
HMM	*YWRIPERFLGDQvT=YGGkLe*		
	Y+R	+ LG+++ + G + +	
Query	538	YFRFAQNTLGNKFLFACGPDAK	558

DKF2phutel_2h3

group: differentiation/development

DKF2phutel_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits
complete cds according to E25 start at Bp 56
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp

Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```

1  GGACCGAGGC  TGCACCGGCA  GAGGCTGCGG  GCGGACGCG  CGGGCCGGCG
51  CAGCCATGGT  GAAGATTAGC  TTCCAGCCCG  CCGTGGCTGG  CATCAAGGGC
101  GACAAGGCTG  ACAAGGCGTC  GGCGTCGGCC  CCTGCGCCCG  CCTCGGCCAC
151  CGAGATCCCTG  CTGACGCCGG  CTAGGGAGGA  GCAGCCCCCA  CAACATCGAT
201  CCAAGAGGGG  GAGCTCAGTG  GGCGGCGTGT  GCTACCTGTC  GATGGGCATG
251  GTCGTGCTGC  TCATGGGCCT  CGTGTTCGCC  TCTGTCTACA  TCTACAGATA
301  CTTCTTTCTT  GCACAGCTGG  CCGAGATAA  CTTCTCCCG  TGTGGTGTGC
351  TGTATGAGGA  CTCCTGTCC  TCCAGGTCC  GGACTCAGAT  GGAGCTGGAA
401  GAGGATGTGA  AATCTACCT  CGACGAGAAC  TACGAGCGCA  TCAACGTGCC
451  TGTGCCCCAG  TTTGGCGGCG  GTGACCCGTC  AGACATCATC  CATGACTTCC
501  AGCGGGGCTC  GACTGCGTAC  CATGATATCT  CCCTGGACAA  GTGCTATGTC
551  ATCGAACTCA  ACACCACCAT  TGTGCTGCCC  CCTCGCAACT  TCTGGGAGCT
601  CCTCATGAAC  GTGAAGAGGG  GGACCTACCT  GCCGCAGACG  TACATCATCC
651  AGGAGGAGAT  GGTGGTCACG  GAGCATGTCA  GTGACAAGGA  GGCCCTGGGG
701  TCCTTCATCT  ACCACCTGTG  CAACGGGAAA  GACACCTACC  GGCTCCGGCG
751  CCGGGCAACG  CGGAGGCGGA  TCAACAAGCG  TGGGGCCAAG  AACTGCAATG
801  CCATCCGCCA  CTTCGAGAAC  ACCTTCGTGG  TGGAGACGCT  CATCTGCGGG
851  GTGGTGTGAG  GCCCTCCTCC  CCCAGAACC  CCTGCCGTGT  TCCTCTTTTC
901  TTTCTTCCAG  CTGCTCTCTG  GCCCTCCTCC  TTCCCTCTGC  TTAGCTTGTA
951  CTTTGGAGCG  GTTCTATAG  AGGTGACATG  TCTCTCCATT  CCTCTCCAAC
1001  CTTGCCCCACC  TCCCTGTACC  AGAGCTGTGA  TCTCTCGGTG  GGGGGCCCAT
1051  CTCTGCTGAC  CTGGGTGTGG  CGGAGGGAGA  GGCGATGCTG  CAAAGTGTTC
1101  TCTGTGTCCC  ACTGTCTTGA  AGCTGGGCCT  GCCAAAGCCT  GGGCCACAG
1151  CTGACCCGGC  AGCCCAAGGG  GAAGGACCGG  TTGGGGGAGC  CGGGCATGTG
1201  AGGCCCTGGG  CAAGGGGATG  GGGCTGTGGG  GGCGGGGCGG  CATGGGCTTC
1251  AGAAGTATCT  GCACAATTAG  AAAAGTCCTC  AGAAGCTTTT  TCTTGGAGGG
1301  TACACTTTCT  TCACTGTCCC  TATTCTTAGA  CCTGGGGCTT  GAGCTGAGGA
1351  TGGGACGATG  TGCCACGGGA  GGGACCCACC  AGAGCACAAG  AGAAGGTGGC
1401  TACCTGGGGG  TGTCCAGGG  ACTCTGTCAG  TGCTTCAGC  CCACCAGCAG
1451  GAGCTTGGAG  TTTGGGGAGT  GGGGATGAGT  CCCTCAAGCA  CAACTGTTCT
1501  CTGAGTGGAA  CCAAAGAAGC  AAGGAGCTAG  GACCCCAAGT  CCTGCCCCCC
1551  AGGAGCACAA  GCAGGGTCCC  CTCAGTCAAG  GCAGTGGGAT  GGGCGGCTGA
1601  GGAACGGGGC  AGGCAAGGTC  ACTGCTCAGT  CACGTCCACG  GGGGACGAGC
1651  CGTGGGTTCT  GCTGAGTAGG  TGGAGCTCAT  TGCTTTCTCC  AAGCTTGGAA
1701  CTGTTTGA  AGATAACACA  GAGGGAAAGG  GAGAGCCACC  TGGTACTTGT
1751  CCACCCCTGCC  TCCTCTGTTC  TGAAATTCCA  TCCCCTCAG  CTTAGGGGAA
1801  TGCACCTTTT  TCCCTTTCTT  TCTCACTTTT  GCATGTTTTT  ACTGATCATT
1851  CGATATGCTA  ACCGTTCTCA  GCCCTGAGCC  TTGGAGAGGA  GGGCTGTAAC
1901  GCCTTCAGTC  AGTCTCTGGG  GATGAAACTC  TTAATGCTT  TGTATATTTT
1951  CTCATATTAGA  TCTCTTTTCA  GAAGTGCTA  TAGAACATA  AAAATCTTTT
2001  ACTTCTGAAA  AAAAAAAAAA  AAAAGGCGG  CCG

```

BLAST Results

Entry B64417 from database EMBL:
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.
Length = 715
Plus Strand HSPs:

Score = 1546 (232.0 bits), Expect = 7.8e-64, P = 7.8e-64
Identities = 310/311 (99%)

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA
library subtraction.
Molecular cloning and characterization of a gene belonging to a novel
multigene family of
integral membrane proteins.

Peptide information for frame 2

ORF from 56 bp to 856 bp; peptide length: 267
Category: strong similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGSSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VVPVQFGGDP ADIIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRFNW ELLMNVKRGY YLPQTYIIQE
201 EMVVEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRRIN KRGAKNCNAI
251 RHFEFTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_2h3, frame 2

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16)., N = 1, Score = 573, P = 1.3e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N =
1, Score = 560, P = 3.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1,
Score = 456, P = 3.3e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16).

Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.3e-55, P = 1.3e-55
Identities = 117/264 (44%), Positives = 172/264 (65%)

```

Query: 1 MVKISFQPAVAGIKGDKADKASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY 60
      MVK+SF A+A + A+K ++ ++L+ P ++P G
Sbjct: 1 MVKVSFNSALA--HKEAANKEEENS-----QVLILPPDAKEPEDVVVPAGHKRAWCWC 51

Query: 61 LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL-----SQVRTQM-- 112
      + G+ +L G++ Y+Y+YF Q + CG+ Y ED LS +Q+++
Sbjct: 52 MCFGLAFMLAGVILGGAYLYKYFAFQ---GGVYFCGIKIEDGLSLPESGAQLKSARYH 108

Query: 113 ELEEDVKIYLDENYERINVPVQFGGDPADIHDFQRLTAYHDISLDKCYVIELNTTI 172
      +E+++I +E+ E I+VVPV+F DPADI+HDF R LTAY D+SLDKCYVI LNT++
Sbjct: 109 TIEQNIQILEEEDVEFISVPVPEFADSDPADIHDFHRLTAYLDLSLDKCYVIPLNTSV 168

Query: 173 VLPFRNFWEMLLMNVKRGTYLPQTYIIQEEMVVEHVSDKEALGSFIYHLCNGKDTYRLRR 232
      V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+R
Sbjct: 169 VMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVQDLGFFFIYRLCRGKETYKLQR 228

Query: 233 RATTRRRINKRGAKNCNAIRHFEFTFVVETLIC 264
      + + I KR A NC IRHFE F +ETLIC
Sbjct: 229 KEAMKGIQKREAVNCRKIRHFEFTFVETLIC 260

```

Pedant information for DKFZphut1_2h3, frame 2

Report for DKFZphut1_2h3.2

[LENGTH] 267
 [MW] 30253.96
 [pI] 8.16
 [HOMOL] SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
 1e-49
 [PROSITE] MYRISTYL 4
 [PROSITE] PRENYLATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 15.36 %

SEQ MVKISFQPAVAGIKGDKADKASASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY
 SEGXXXXXXXXXXXXXXXXX.....
 PRD cccccccchhhhhhhhhhhhhhhcccccceccccccccccccccccccccchh
 MEMMMM

SEQ LSGMVLVLLMGLVFASVYIYRYFFLAQLARDNFFRCGLVYEDSLSSQVRTQMELEEDVKI
 SEG ..XXXXXXXXXXXXX.....
 PRD hhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhccceccccccccchhhhhhhhhhh
 MEM MMMMMMMMMMMMMMMMMMMMM

SEQ YLDENYERINVPVPQFGGDPADIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNFV
 SEG
 PRD hhccccccccccccccccchhhhhhhhhhhhhccceccccccccchhh
 MEM

SEQ ELLMNVKRGTYLPQTYIIQEEMVTEHVSDEALGSFIYHLCNGKDTYRLRRRATRRRIN
 SEGXXXXXXXXXXXXX.....
 PRD hhhhhccccccccceehhhhhhhccccchhhhhheccccchhhhhhhhhhhhh
 MEM

SEQ KRGAKNCNAIRHFENTFVVETLICGVV
 SEG xx.....
 PRD hhhccccccccchhhhhhecccc
 MEM

Prosites for DKFZphut1_2h3.2

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	50->54	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PENYLATION	PDOC00266

(No Pfam data available for DKFZphut1_2h3.2)

DKFZphmcfl_1a11

group: transmembrane protein

DKFZphmcfl_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcfl_1a11 encodes a novel 393 amino acid protein, with
similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits

potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGGCCAG CCCCGAAGA GCCGCTCAG CCGGGGGGAG TTGCTCGGAC
51  TCAAACTGCC AGTCTCTGTG CGACCGCGCT GGGTCGGAAG TGAGCAGGCT
101 GAGGCCACCA TGGAGCAGTG TGCCTGCCGTG GAGAGAGAGC TGGACAAGGT
151 CCTGCAGAAAG TTCCTGACCT ACGGGCAGCA CTGTGAGCGG AGCCTGGAGG
201 AGTGTGTGCA CTACGTGGGC CAGCTGCCGG CTGAGCTGGC CAGCGCAGCC
251 CTCAGGGGGA CCCCTCTCTC AGCCACCCCTC TCTCTGGTGA TGTACAGTG
301 CTGCCGGAAG ATCAAGATA CGGTGCAGAA ACTGGCTTCG GACCATAAGG
351 ACATTACAGC CAGTGTATCC CGAGTGGGCA AAGCCATTGA CAGGAACCTC
401 GACTCTGAGA TCTGTGGTGT TGTGTCTGAT GCGGTGTGGG ACGCGCGGGA
451 ACAGCAGCAG CAGATCCTGC AGATGGCCAT CGTGAACAC CTGTATCAGC
501 AGGCGATGCT CAGCGTGGCC GAGGAGCTGT GCCAGGAATC AACGCTGAAT
551 GTGGACTTGG ATTTCAAGCA GCCTTTCCTA GAGTTGAATC GAATCTTGA
601 AGCCCTGCAC GAACAAGACC TGGGTCTTGC GTTGAATGG GCCGTCTCCC
651 ACAGGCGGCG CCGTCTGGAA CTCAACAGCT CCCTGGAGTT CAAGCTGCAC
701 CGACTGCACT TCATCCGCTT CTTGGCAGGA GGCCCCGCGA AGCAGCTGGA
751 GGCCCTCAGC TATGCTCGGC ACTTCCAGCC CTTTGCTCGG CTGCACCAGC
801 GGGAGATCCA GGTGATGATG GGCAGCCTGG TGTACCTCGG GCTGGGCTTG
851 GAGAAGTCAC CCACTGCGCA CCTGTGGAG AGCAGCCACT GGGCAGAGAT
901 CTGTGAGACC TTTACCGGGG ACGCCTGTTT CCTGTGGGG CTTTCTGTGG
951 AGTCCCCCTT TAGCGTCAGC TTTGCCTCTG GCTGTGTGGC GCTGCCTGTG
1001 TTGATGAACA TCAAGGCTGT GATTGAGCAG CGGCAGTGCA CTGGGGTCTG
1051 GAATCACAAG GACGAGTTAC CGATTGAGAT TGAACAGGCC ATGAAGTGCT
1101 GGTACCACCT CGTGTTCGCT TGCCCCATCC TCCGCCAGCA GACGTCAGAT
1151 TCCAACCCCT CCATCAAGCT CATCTGTGGC CATGTTATCT CCCGAGATGC
1201 ACTCAATAAG CTCATTAATG GAGGAAAGCT GAAGTGTCCC TACTGTCCCA
1251 TGGAGCAGAA CCCGGCAGAT GGGAAACGCA TCATATTCTG ATTCTACCT
1301 GGAAGGAATT TTGTTGAAAG GGGTTTTCAC CTGTGAGCCT TGGTCTGTCT
1351 CGGTAGGGTG GTCAACTTCA GTGGACTGTG GTTGGTTTCA GAGCGCCTGG
1401 CTGAGGAGTT CCACTGAGGG GAGCACTGGA GCAGCCCTTT GGCAGAGGCT
1451 GAGGAGGGAG ATGGACAGC CCACGCTTGG CACCTGGCTC CATGGCATAA
1501 GGAAGGGGAG ATGCTGGCCT CTGTGCTCCT GCTGTCTTTT CCTGTTTCTG
1551 TTTGCGTTTG ACTTAGTAGC AACCGACAGA GTGGCAAGGG ATTTGGTCTT
1601 CAGCAGTAGA CATCTTCCA CCCCTGCCCT CAGCCAAGTC TCTGTCTGCC
1651 ATGCCAATGC TATGTCACC CTTGCCCTCC GGCCAAGAG TGTCCAGCGG
1701 TGGCCACCTT TTCTCTCCA CTACAGCCTC AACAGTATGT ACCATCTCCC
1751 ACTGTAAATA GTCCAGTTA GAACGGAATG CCGTTGTTTT ATAACTTTGA
1801 ACAATGTAA AAAAAAAAAA

```

BLAST Results

Entry HS579359 from database EMBL:
human STS WI-6350.
Score = 1027, P = 9.9e-40, identities = 207/209

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQHCERSLE ELLHYVGQLR AELASAAQOG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSSVSRVG KAIDRNFDSE
101 ICGVVSDAVW DAREQQQIIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETFTF DACSLGLSV ESPLSVSFAS GCVALPVLMM
301 IKAVIEQRQC TGVWNHKKDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLCIGHVI SRDALNKLIN GGLKCPYCP MEQNPADGKR IIF
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_lal1, frame 2

TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3. Length = 398

HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query: 252 YCHLLDSSHWAIEICETFTRDACSLGLSVESPLSVSFASGCVALPVLMMNIKAVIEQRQCT 311
 Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ + +++++
 Sbjct: 258 YIDVLDLD-WKSLELLFVREFCAALGMSLESPLDIVVNAGAIAPILLKMSSIMKKKHT 316

Query: 312 GVNHNKDELPIEIELGMKCWYHSVFACPIRQQTSDSNPPIKLCIGHVISRDALNKLING 371
 W + ELP+EI L +HSVF CP+ ++Q ++ NPP+ + CGHVI ++L +L
 Sbjct: 317 --WTSQGELPVEIFLPSSYHFSVFTCPVSKEQATEENPPMMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKCPYCPMEQNPADGKRIF 393
 G + KCPYCP E AD R+ F
 Sbjct: 375 GSQRFKCPYCPNENVAADAIRVF 398

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query: 22 GQHCERSLEELLHYVGQLRAELASAAQGTPLSATLSLVMSQCCRKIKD TVQKLASDHKD 81
 G C L EL + + + L+ P ++ LV C K + L K
 Sbjct: 15 GNKCLAKLNL---ESILKDAKKSCLKD-PTTSMKELVA--CSEKTTQVFDLKRTEKK 67

Query: 82 IHSSSVRVGKAIDRNFDSEICGVVSDAVWDAREQQQIILQMAIVEHLYQQGMLSVAEELC 141
 H+S++R GK +++ F+ ++ + ++++++ + A+ H ++Q + +A C
 Sbjct: 68 FHTSLNRFGKTLEKKFNFDLEDIKLHSSFESKKRE---IDTALSLHFFRQGDVELAHLFC 124

Query: 142 QESTLNVDLDFKQPFLELNRIEALHEQDLGPALEWAVSHRQRLLELNSLEFKLHRLHF 201
 +E+ + + F L I++ ++DL +EWA R L SSLE+ L +
 Sbjct: 125 KEAGIEEPSSELHVFTLLKSIVQGIRDKOLKLPIEWASQCRGYLERKGSSELYTLQKYRL 184

Query: 202 IRLLAGGPAKQL-EALSYAR-HFQPFARLHQREIQVMMGSLVY 242
 + K + A+ Y R + F + H +IQ M +L +

Pedant information for DKFZphmcf1_1a11, frame 2

Report for DKFZphmcf1.1a11.2

[illegible]

Prosite for DKFZphmcf1 1a11.2

PS00001	189->193	ASN_GLYCOSYLATION	PDOC00001
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	27->36	TYR_PHOSPHO_SITE	PDOC00007
PS00007	244->253	TYR_PHOSPHO_SITE	PDOC00007
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	50->56	MYRISTYL	PDOC00008
PS00009	387->391	AMIDATION	PDOC00009
PS00013	282->293	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1a11.2)

DKF2phmcf1_1c23

group: mammary carcinoma derived

DKF2phmcf1_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
1  AACTGGCCCC CTCCCCACC CCCTGCCCCT GAGGAGCAGG ACCTGTCCAT
51 GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCCTTCAGGC TCCCCAGAGC TTGTCAAGTC CCGGCTGCT
151 TCGTCTCTCT CAGCTACTGC TTTCAGATT CAGCCCCCGG GTAGCCAGAG
201 CCCTCTCTCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCTCCCA GGGAGGACGT AGGTGCGCCC CTGGTCACGC CCTCGCTCCT
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCACCCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
451 GGGCGGGCCA GCCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGGGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCTGAG GCAGAGCCAC GGCCCTCCCA GTCCCTTGCC
601 TCAACGGCCA GTTTCATCTT CTCCAAGGCG TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCGTG TCCCCTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGGC CCAAGAAGAG
751 TCACCTAAGG CTCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
801 CCCACCCGCG TCCCCAGTT ACCCTCGAGC TGAGCCCTTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAAATGGAG GTGTCTCTCA GCTGGTGGGC CCAGAGGAGA AGATGGGCTC
951 CCGGGGCTCA GACTCAGAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGACAGCTC
1051 CCAAGGACGA GAGGATACAG CAGACACAAC CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCCT TCGATACCTA TGCAAGCCTG GTGTGTCTCC
1151 TGTCTCTAGA GTCATCTGCG GCTCATGCCT TTCCCGAAT GGGTTACCTT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCTGTA CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTTTTCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTTGAAAA GTGGCTGCAT
1401 GGGCATCCTC CAGGGCCAGG GAAGTTGCAT TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCCTCTGCG CCAGTCCAGT TTACTCCTCT TGGTCTCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCACAGGCG TTCTTAGGCG TGGGGGCAGG TGTGGGGCTG
1601 TGGAAATCCA AAGCACAAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACTCACCA ACCACCTCCG TGCTTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCCTCCTT
1801 GACCACACTG GGGTGGGTGG GCCCCCAGCA CTTCAGAGGC AGGAGCCTTT
1851 GGGCTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCTT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCTTCA TTTCGGTGCA TGTCTTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGGAAAGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TTAGGGCAG GCTGGGTGGT GGTCCCTTAG GTATCAGCCT
2151 CTCTTACTGT ACTCTCCGGG AATGTTAACC TTCTATTITT CAGCCTGTGC
2201 CACCTGTCTA GGAAGCTGG CTTCGCCATT GGCCCTGTGT GGTCCACAGC
2251 AGCGTGGCTG CCCCCCAGGG CCACCGCTTC TTCTTTGATC CTCTTTCCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAGCTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCACTG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTCTGCCTC
2601 AGAGCTTCTG ATCTACCCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTTCCC TCTCCTCCT CCAACATTAC TGGAACTCTA
```

```

2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AACTATATAT ATACATATAT ATATTTCTTT
2951 AAATTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAACTGTG TTTTATTAA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311
 Category: putative protein
 Classification: unset

```

1 MADFPPEEAF FFSVASPEPA GPSGSPPELV SPAASSSSAT ALOIQPPGSP
51 DPPPPAPPAPA PASSAPGHVA KLPOKEPVGC SKGGGPPRED VGAPLVTPLS
101 LQMVRLRSVG APGGAPTAL GPSAPQKPLR RALSGRASP PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERVSPETQA DLQRNLVAEL RSISEQRPPQ APKKSPPKAP FVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGVL QLVGPEEKMG
301 LPGSDSQKEL A

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_lc23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize
 Length = 1,188

HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15
 Identities = 81/269 (30%), Positives = 115/269 (42%)

```

Query:      5 PPPEEAFFS---VASPEPAGPSGSPPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
            PPP  S  V SP P P SP  PA +SS ++ PP +P PPP  +
Sbjct:    598 PPPPAPVASPPPPVKSPPTPVASPP---PPAPVASSPPPMKSPPTPVSSPPPEKS 654

Query:      56 PPAPAPASSAPGHVAKLPQKEPVGC SKGGGPPREDVGCAPLVTPLSLLQMVRLRSVGAPGGA 115
            PP P PA S P  + P  P  K  PP ++ P + PS  +  P
Sbjct:    655 PPPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKTPPSTPSKP 711

Query:     116 PTPALGPSAPQKPLRRA-LSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQFNGPPE 174
            P+  PS P++P+  + ++SP PAP S  +LA  S + + PP
Sbjct:    712 PSSPEKPSPPKEPVSSPPTPKSSPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query:     175 AEPRPQSPASTASFI FSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPPQAPK 233
            PP +P  +S  +Q+ P +P++ L  V+  + + PP AP
Sbjct:    772 PLSSPPAPQVKSS-----PPPVQVSSPPAPKSSPPLAP--VSSPPQVKTSPPPPAPL 823

Query:     234 KSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPPTNGLP 273
            SP  P  + P V V PPP  S P  P+++PP  P
Sbjct:    824 SSPLAPK-SSPPHVVSPPPPVVKSSPPAPVSSPPLTPK 864

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14

```

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPELVSSPAASS---SSATALQIQPGSPDPPFPAP---PAPAPASSAPGHV 69
 P P G P SP + PAAS+ S T + P P+P P P P P P +P
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSPHTPPDVSPLEPLPEPSVPAPAPMPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGPSAPQKP 128
 +P PV G S P V P + +V+L AP G+P P + ++P P
 Sbjct: 469 DYVPTTPPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTSPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188
 + G SP P P S + +K+ A G + P PPE P PP AS
 Sbjct: 529 I-----GSPSP-PPPVSVSPPPPVKSPPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSKGSRKLQLERPVS PETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVARKPS- 247
 + S L P P ++ VA + PP P SP P PVA P
 Sbjct: 578 PVKSPPPPTLVASPP--PPVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVASSPPP 635

Query: 248 VGVPPP---ASPSYPRAEPLTAPPTNGLPHTQD 277
 + PPP +SP P P PP P ++
 Sbjct: 636 MKSPPPTPVSSPPPEKSPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPELV---SSP--AASSSSATALQIQPGSP-DPPAPAPAPASSAPGHVA 70
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+
 Sbjct: 817 SPPPA-PLSSPPLAPKSSPPHVVSPPPPVKSPPPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ---KEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGPSAPQ 126
 P+ P + PP E +P TP L ++S P +P + P +
 Sbjct: 873 SPPEVVKPSTPPAPTIVISPPSEPKSPPTPVSLPPPIVKSSPPFAMVSSPMTPKSSP 932

Query: 127 KPLRRAL---SGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P
 Sbjct: 933 PPVVSSPPPTVKSSPPFAPVSSPPATP--KSSPPFAPVNL---P--PPEVKSPPPTP 984

Query: 184 ASTASFIFSKGSRKLQLERPVS PETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVA 243
 S+ + P PE ++ V+ + PP AP SP PPPV
 Sbjct: 985 VSSPPFAPKSSPPFAPMSSPPPEVKSPPFAPVSSPPPVKSPPFAPVSSP--PPPVK 1042

Query: 244 RKPS---VGVPPPASPSPYRAEPLTAPP 268
 P V PPP S P P++PP
 Sbjct: 1043 SPPFAPVSSPPPVKSPPFAPISSP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPELVSSPAAS-SSSATALQIQPGSPDPPFAPAPAPAS 63
 PPP S PE + P P + P + T+++ PP PP P+P
 Sbjct: 639 PPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPPPTLIPSPPP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGPS 123
 P K P K PP+E V +P TP V +P PTP P
 Sbjct: 699 QEKPTPTSTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPFAPVSSP--PPTPVSSPP 753

Query: 124 APQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
 A P+ S ++SP PAP S A ++K+ + + P PP + PP +P
 Sbjct: 754 A-LAPVSSPPSVKSSPPFAPLSSPPFAPQVKS---SPPPVQVSSP--PPAPKSSPLAP 806

Query: 184 ASTASFIFSKGSRKLQLERP-VSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPV 242
 S+ + L P ++P++ +V+ + + PP AP SP P
 Sbjct: 807 VSSPPQVEKTSPPFAPLSSPPLAPKSSPP--HVVVSSPPPVKSPPFAPVSSPPLTPKP 864

Query: 243 ARKPS-VGVPPP---PASPSYPR-----AEPLTAPP 268
 A P+ V PP P++P P +EP ++PP
 Sbjct: 865 ASPFAHVSSPEVVKPSTPPAPTIVISPPSEPKSSPP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPEL-VSSP--AASSSSATALQIQPG--SPDPPFAP--- 56
 PPP A S P P S P + VSSP A SS A PP PPPAP
 Sbjct: 768 PPP--APLSSPPFAPQVKSPPPVQVSSPPFAPKSSPLAPVSSPPQVEKTSPPFAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAP 116
 P AP SS P V P PV S PP V +P +TP V +P
 Sbjct: 826 PPLAPKSSPPHVVSPP--PVVKSS---PPFAPVSSPPLTPKPFAPFA--HVSSPPEVV 878

Query: 117 TPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQPN--- 169
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPAPTTVISPPSEPKSSPPPTVSLPPIVKSSPPPPAMVSSPPMTPKSSPPPVV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQRP 228
+ PP + PP +P S+ + P PE ++ V+ + P

Sbjct: 938 SSPPPTVKSSPPPPAPVSSPPATPKSSPPPPAPVNL-PPEVKSSPPPTVSSPPPPAPKSSP 996

Query: 229 PQAPKKSPPKAPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPEVKSSPPPPAPVSSPPPPVKSSPPPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGPSGSELVSSPAASSSATALQIQPP---GSPDPP---PA 55
D+ PP V P S SP+ V PAAS+ + ++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTS 524

Query: 56 PPAPAPASSAPGHVAKL---PQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGA 111
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPPVSVSSPPPPVKSSPPPPAPVGSPPPEKSPPPAPVASPPPPVKSSPP 584

Query: 112 PG--GAPTALGPSAPQKPLRRA---LSGRASVPAPSSGLHAAVRLKACSLAASEGLSS 166
P +P P + P P+ + P P S A V + + +

Sbjct: 585 PTLVASPPPPVKSSPPPPAPVASPPPPVKSSPPPTPVASPPPPAPVASSPPPMKSPPPPTP 644

Query: 167 AQPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQ 226
PPE P PP PA + + ++ PE L+ + +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYTPPTS VKSSPPPEKSLP-PPTLIPSPPPQEK 702

Query: 227 RPPQAPKKSPPKAPP-PVARKPSVGVPPASPSPYPRAEPLTAPP 268
PP P K P +P P K V PP S P P+++PP

Sbjct: 703 TPPSTPSKPPSSPEKSPPEKSPVSSPPQTPKSSPPPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPEEAEFFSVASPEPAGP---SGSELVSSPAASSSATALQIQPGSP--DPPAP-- 56
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPAPVSSPPPLTKPASPPAHVSSPPEVK-PSTPPAPTIV--ISPPSEPKSSPPPTVPS 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGA 115
P P SS P + P P P PP V +P P++ V +P

Sbjct: 907 LPPPIVKSSPPPPAMVSSPPMTPKS-----SPPVVVSSP--PPTVKSSPPPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNPPEA 175
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 PKSSPPPPAPVNLPPPEV---KSSPPPTPVSSPPAPK---SSPPPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQRPQAPKKS 235
+ PP +P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPPPAPVSSPPPPVKSSPPPPAPVSSP-PPPVKSSPPPPAPVSSPPPPVKSSPPPPAPISS 1068

Query: 236 PKAPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P--PPPVKSSPPPPAPVSSPPPPVKSSPPPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPGSPDPPAP---PAPAPASSAPGHV 69
P P G P SP + PAAS+ S T + P +P P P P P P +P

Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSPTTPDVSPPELPEPSVPVAPAPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGPPREDVGAPLVTPSLLQMVRLRSVGA 128
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNPPEAEPRPPQSPASTAS 188
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVSSPPPPVKSSPPPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSKGSRKLQLERPV---SPETQADLQRLVAELRS-----ISEQRPPQA-----PK 233
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSSPPPTLVASPPPPVKSSPPPPAPVA-SPPPVKSSPPPTPVASPPPPAPVASSPPPM 636

Query: 234 KSPKAPPPVARKP---SVGVPPASPSPYPRAEPLTAPPTN 270
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPTPVSSPPPEKSPPPPPAKSTPPPEEYTPPTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFVSASPEPAGSGSPSELVSSPAASSSSATALQIQQSPDPPAPAPAPASS 64
 PP S S + P +P + P SS A+ PP +P +PP P SS
 Sbjct: 883 PPAPTVISPPSEPKSSPPPTVSLPPPIVKSSPPPMVSSPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPG--GAPTPALGP 122
 P V P PV PP +P P L ++S P +P PA
 Sbjct: 940 PPPTVKSSPPAPVS-----SPPATPKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRFP 180
 S P P+ ++ P PAP S V+ S +SS P PP + PP
 Sbjct: 995 SPPAPMSSPPPEVKSSPPPAFVSSPPPVK----SPPPAFVSS--P--PPPVKSSPP 1046

Query: 181 QSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKSPPKAP 240
 +P S+ + P P ++ V+ + PP AP SP PP
 Sbjct: 1047 PAPVSSPPPVKSSPPPAPISSP-PPPVKSSPPPAFVSSPPPVKSSPPPAFVSSP--PP 1103

Query: 241 PVARKPS---VGVPASP---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283
 P+ P V PPPA PS P P+++PP P + ++ L
 Sbjct: 1104 PPKSPPPAFVSSPPPAFVKPPLPPPAFVSSPPPVTPAPPKKEEQL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFVSASPEPAGSGSPSELVSSPAASSSSATALQIQQP---GSPDPP---PA 55
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +
 Sbjct: 469 DVPPTTP---VPGKSPATSPSPQ-VQFPAASTPPSLVLKSPQAPVGSPPPVKTT 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGA 115
 PPAP + S P V+ + PV PP VG+P P V+P
 Sbjct: 525 PPAPIGSPSPFPVSVVSPPPPVKSP---PPPAFVGS--PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175
 P P P P ++ P PAP + V+ S ++S P P +
 Sbjct: 576 PPPVKSPPPTLVASPPPVKSSPPPAFVASPPPVK---SPPPTPVASPPPAFVAS 631

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
 P P +SP K P P S+ PP+
 Sbjct: 632 SPPMKSPPPPTPVSSPPPEKSP--PPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPP 689

Query: 236 PK---APPPVARK--PSVGVPASPSPYRA--EPLTAPP 268
 P +PPP + PS PP+SP P EP+++PP
 Sbjct: 690 PTLIPSPPPQEKPTPSTPSKPPSSPEKSPPKPEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPEEAFVSASPE-PAGPSGSPSELVSSPAASSSSATALQIQQSPDPP-PAPPAP 59
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P
 Sbjct: 496 ASTPPP--SLVKLSPQAPVGSPPPVKTTSPAPIGSPSPFPVSVVSPPPPVKSPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGAPGAPTA 119
 AP S P P PV PP + P +S V+ AP +P P
 Sbjct: 554 APVGSPPPEKSPPPPAFVASP---PPVKSPPPTLVASPPPVKSSPPPAFVASPPPP 610

Query: 120 LGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPR 178
 + P P+ + P PAP + ++ +S P PP A+
 Sbjct: 611 VKSPPPPTPVA-----SPPPAFVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKST 664

Query: 179 PP--QSPASTASFIFSKGSRKLQLERP---SPETQADLQRNLVAELRSISEQRPPQAPK 233
 PP + P S S K L P SP Q S ++P +P
 Sbjct: 665 PPPEEYPTPTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPSTPSKPPSSPEK--SPP 721

Query: 234 KSPKAPPPVARKPSVGVPASPSPYRAEPLTAPP 268
 K P + PP K S PPPA S P P+++PP
 Sbjct: 722 KEPVSSPPQTKSS---PPPAFVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGSGSPE-LVSSPAASSSS---ATALQIQQSPDPPPP-- 54
 PPPE++ VASP P S P LV+SP S A PP PPP
 Sbjct: 560 PPPEKSPPPAPVASPPPVKSPPPPTLVASPPPVKSSPPPAFVASPPPVKSPPPPT 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVG---SKGGPPREDVGAPLVTPSLLQMVRLRS 108
 +PP PAP +S+P + P PV K PP P ++S
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKS 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSA 167
 P + P P L PS P P + + ++P PSS + + S SS
 Sbjct: 680 SPPPEKSLPPPTLIPSP--PQEKPTPSTPSKPPSSPEKSPPKPEPVSSPPQTKSSP 736

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQR 227
P P P SP + A + S S K P + P + + + +
Sbjct: 737 PPAVSSPPPTFVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPPVQVSS 793

Query: 228 PPQAPKSKPAPPEVARKPSVGVPPPASPSYPRAEPLTAPP 268
PP APK SP P+A P V PP + P PL++PP
Sbjct: 794 PPPAPKSSP---PLA--P-VSSPPQVEKTSPPAPLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09
Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFFSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQPPGSPDPPPP-APPAPA 60
PPP + + + P P G PS P +VS P S P GSP PP +PP PA
Sbjct: 517 PPPVK---TTSPPAPIGSPSPPPVSVSPPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP---LVTPSLLQMVRLRSVGAPGG 114
P+S P V P V PP V+P + +P V AP
Sbjct: 571 PVASPPPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASVPVAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171
+ P + P P+ SP P P S+ S+ +S + P
Sbjct: 631 SSPPPMKSPPPPTFVSSPPPEKSPPPPPAKSTPPPEEYPTPTPSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQA 231
PP P PP T SK P SPE + + V+ + PP A
Sbjct: 689 PTLIPSPPPQEKPTPPSTPSKP-----PSSPEKPS-PKEPVSSPPQTPKSSPPPA 739

Query: 232 PKKSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268
P SP P PV+ P++ PP+ S P PL++PP
Sbjct: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPAPLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08
Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFFSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQPPGSPDPPPPAPPAPA 60
A P P SPEP PS P P + S A PP P P +PPA +
Sbjct: 427 ASAPMPSPHTPPDVSPELPEPSVPAPAPMPMPTPHSPPADDDYVPTTPVVGKSPPTS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGATP-- 118
+ A P V S PP+ VG+P P V+ S AP G+P+P
Sbjct: 487 PSPQVQPPAATFPPSLVKLS---PPQAPVGSF--PPF---VKTSPPAPIGSPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
+ P P K P A G SP P S A S + + PP
Sbjct: 537 PVSVPSPPPVKSPPPAPVG--SPPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPP 594

Query: 175 AEP RPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKK 234
+ PP +P + + P P A + + PP P+K
Sbjct: 595 VKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPPTNGLP 273
SP PPP P PP P+ P + + PP LP
Sbjct: 654 SPPPPPPAKSTP---PPEEYPTPTPSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08
Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFFSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQPPGSP--DPPAP---PAP 59
PPP V+SP P P SP P SS ++ PP +P PP P P P
Sbjct: 916 PPPA---MVSSP-PMTKSSPP---PVVSSPPPTVKSSPPAPVSSPPATPKSSPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGATPA 119
AP + P V P PV S P AP+ +P + V+ AP +P P
Sbjct: 967 APVNLPPPEVKSSPPPTPV-SPPAPKSSPPAPMSSPPPE-VKSPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPVAPSSGLHAAVRLKACSLAASEG---LSSAQPNGPPEA 175
+ P P+ ++ P PAP S V+ S + S P P +
Sbjct: 1025 VKSPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKS 235
P P +SP A S ++ P P A + A ++ S PP AP S
Sbjct: 1085 PPPPVKSPPPAPV---SSPPPIKSPPP---APVSSPPAPVKPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYRAEPLTAPP 268
P P +K +PPPA S P + PP
Sbjct: 1136 PPPVVTAPPKKEEQSLPPFAESQPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06
Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFPPEEAFSSVASPEP-AGPSGSELVSSPAASSSSATA-LQIOPPGSP--DPPP---A 55
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +
 Sbjct: 970 NLPPPEVK--SSPPPTPVSSPPPAKSSPPAPMSSPPPEVKSPPPAPVSSPPPPVK 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115
 PP PAP SS P V P PV PP + P S V+ AP +
 Sbjct: 1028 PPPPAPVSSPPPPVKSPPPAPVSSPP---PPVKSPPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 116 PTFALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174
 P P + P P+ ++ P PAP S A +K SL +SS P PP
 Sbjct: 1085 PPPPVKSPPPPAPVSSPPPIKSPPPAPVSSPPAP-VKPPSLPPAPVSS--P--PPV 1139

Query: 175 AEP RPQ 181
 P PP+
 Sbjct: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPEEAFSSVASPEPAGP-SGSELVSSP---AASSSSATALQIOPPGSP--DPPP 54
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1001 MSSPPPE----VKSPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPP 1055

Query: 55 ---APPAPAPASSAPGHVAKLPQKEPVGCSK---GGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP SS P V P PV PP V +P P +
 Sbjct: 1056 PVKSPPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125
 V +P AP P+L P AP
 Sbjct: 1114 VSSPPAPVKKPPLPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFSS---VASPEPAGP-SGSELVSSP---AASSSSATALQIOPPGSP--DPPP 54
 PPP S V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1060 PPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPPPIKSPPPAPVSSPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108
 AP P PAP SS P V P K+ + PP E P +L +
 Sbjct: 1120 APVKPPSLPPAPVSSPPPVVTPAPPKKE---EQSLPPPAESQPPSFNDIILPIMANK 1176

Query: 109 VGAP 112
 +P
 Sbjct: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQPNG 171
 G PTP GP + P + A S +P+P+ + L S + A + P+
 Sbjct: 408 GYPTPGGGPPSSPVGKPAAS---APMPSHTPPDVSEPLEPEPSVPAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFIKSGSRKLQLERPVSFETQ---ADLQRNLVAELRSISEQR 227
 PP + PP P S + S ++Q +P + Q + + +
 Sbjct: 465 PPADDYVPPTPPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 228 PPQAPKKSFKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 PP AP SP PPPV SV PPP S P P+ +PP
 Sbjct: 525 PP-APIGSPSPPPV---SVVSPPPPVKSPPPAPVGSPP 560

Pedant information for DKFZphmcf1_lc23, frame 1

Report for DKFZphmcf1_lc23.1

[LENGTH]	311
[MW]	31534.58
[pI]	9.48
[KW]	All Alpha
[KW]	LOW_COMPLEXITY 38.59 %

SEQ MADFPPEEAFSSVASPEPAGPSGSELVSSPAASSSSATALQIOPPGSPDPPPPAPPAPA
 SEGXX
 PRD ccc

SEQ PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTAL
 SEG xxxxxx.....XXXXXXXXXXXX

```
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccc
SEQ      GPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAAEGLSSAQNGPPEAEPRPP
SEG      xxxxxx.....xxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      QSPASTASFIFSGSRKLQLLRVPSPETQADLQRNLVAELRSISEQRPPQAPKSPKAPP
SEG      xxxxx.....xxxxxxxxxxxxxxxx
PRD      cccccceeeccccchhhhhccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      PVARKPSVGVPFPASPSYPRAEPLTAPPTNGLPHTQDRTKRELAENGVLQLVGPEEKMG
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccecccccccccc

SEQ      LPGSDSQKELA
SEG      .....
PRD      ccccccccccc
```

(No Prosite data available for DKFZphmcf1_1c23.1)

(No Pfam data available for DKFZphmcf1_1c23.1)

DKFZphmcf1_1e15

group: transmembrane protein

DKFZphmcf1_1e15 encodes a novel 454 amino acid protein with similarity to *C. elegans* proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER
membrane regions: 9

complete cDNA, complete cds, EST hits
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp

Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGCAGCGC  CCGGGCTGAG  CGACAGCAAG  TGCAGCGGGC  TCCTACCCCG
51  GGTGAGGGGT  GGCCTCCGCG  TGGGATCGTG  CCCTCTTCAG  CCCGCTCCTG
101  TCCCGACAT  CACGTGTATT  CCGCACGTCC  CCTCCGCGCT  GTGTGCTAC
151  TGAGACGGGG  AGGCGTGACA  GGGCCCGGGT  CCCTTCTCAG  TGGTGCTCTG
201  TGCTTCAGGG  CAAGCTCCCC  GTCTCCGGGC  GCACTTCCTT  CGCTGTGTGT
251  CGGTCCATCC  TCCTTTCTCC  AGCCTCCTCC  CCTCGCAGGT  GGGATCGTGG
301  GTGGGACCGG  AGCGCGGGCG  GCGCGGGCCC  CCGGGGACCA  TGGCCGGGTC
351  CGACACCGCG  CCCTTCCTCA  GCCAGGCGGA  TGACCCGGAC  GACGGGGCAG
401  TGCTTGGCAC  CCGGGGGTTG  CCAGGGTCCA  CGGGGAACCC  GAAGTCCGAG
451  GAGCCCGAGG  TCCCGGACCA  GGAGGGGCTG  CAGCGCATCA  CCGGCCGTGC
501  TCCCGGCCGT  TCGGCTCTCA  TAGTGGCGGT  GCTGTGCTAC  ATCAATCTCC
551  TGAACATCAT  GGACCGCTTC  ACCGTGGCTG  TGTTCATCTC  CAGTTACATG
601  GTGTTGGCAC  CTGTGTTTGG  CTACCTGGGT  GACAGGTACA  ATCGGAAGTA
651  TCTCATGTGC  GGGGGCATTG  CCTTCTGGTC  CCTGGTGACA  CTGGGGTCAT
701  CCTTCATCCC  CGGAGAGCAT  TTCTGGCTGC  TCCTCCTGAC  CCGGGGCTG
751  GTGGGGGTGC  GGGAGGCCAG  TTATTCCACC  ATCGCGCCCA  CTCTCATTGC
801  CGACCTCTTT  GTGGCCGACC  AGCGGAGCCG  GATGCTCAGC  ATCTTCTACT
851  TTGCATTTC  GGTGGGCAGT  GGTCTGGGCT  ACATTGCAGG  CTCCAAAGTG
901  AAGGATATGG  CTGGAGACTG  GCACTGGGCT  CTGAGGGTGA  CACCGGGTCT
951  AGGAGTGGTG  GCCGTTCTGC  TGCTGTCTCT  GGTAGTGGCG  GAGCCGCCAA
1001  GGGGAGCCGT  GGAGCGCCAC  TCAGATTTCG  CACCCCTGAA  CCCCACCTCG
1051  TGGTGGGCAG  ATCTGAGGGC  TCTGGCAAGA  AATCTCATCT  TTGGACTCAT
1101  CACCTGCCTG  ACCGGAGTCC  TGGGTGTGGG  CCTGGGTGTG  GAGATCAGCC
1151  GCCGGCTCCG  CCACTCCAAC  CCGGGGCTG  ATCCCTGGT  CTGTGCCACT
1201  GGCTCCTG  GCTCTGCACC  CTTCTCTTTC  CTGTCCCTTG  CCTGCGCCGG
1251  TGGTAGCATC  GTGGCCACTT  ATATTTTCAT  CTTTATTGGA  GAGACCCCTC
1301  TGTCATGAA  CTGGGCCATC  GTGGCCGACA  TTCTGCTGTA  CGTGGTGATC
1351  CCTACCCGAC  GCTCCACCGC  CGAGGCCCTT  CAGATCGTGC  TGTCCCACTT
1401  GCTGGGTGAT  GCTGGGAGCC  CCTACCTCAT  TGGCTGATC  TGTACCCGCC
1451  TGCGCCGGAA  CTGGCCCCCC  TCCTTCTTGT  CCGAGTTCCG  GGCTCTGCAG
1501  TTCTCGTCA  TGCTCTGCGC  GTTGTGTGGG  GCACTGGGCG  GCGCAGCCTT
1551  CCTGGGCACC  GCCATCTTCA  TTGAGGCCGA  CCGCCGGCGG  GCACAGCTGC
1601  ACGTGCAGGG  CCGTCTGCAC  GAAGCAGGGT  CCACAGACGA  CCGGATTGTG
1651  GTGCCCCAGC  GGGGCCGCTC  CACCCGCGTG  CCGTGGCCA  GTGTGCTCAT
1701  CTGAGAGGCT  GCCGCTCACC  TACCTGCACA  TCTGCCACAG  CTGGCCCTGG
1751  GCCCACCCCA  CGAAGGGCCT  GGGCCTAACC  CCTTGGCCTG  GCCCAGCTTC
1801  CAGAGGGACC  CTGGGCCGTG  TGCCAGCTCC  CAGACACTAC  ATGGGTAGCT
1851  CAGGGGAGGA  GGTGGGGGTC  CAGGAGGGGG  ATCCCTCTCC  ACAGGGGACG
1901  CCCCAGGGC  TCGGTGCTAT  TTGTAACGGA  ATAAATTTG  TAGCCAGAAA
1951  AAAAAAA

```

BLAST Results

Entry E12646 from database EMBL:
cDNA encoding cell growth inhibiting factor.
Score = 3046, P = 2.2e-131, identities = 640/659

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454
 Category: similarity to known protein

```

1 MAGSDTAPFL SQADDDDDGP VPGTPGLPGS TGNPKSEEP E VPDQEGLRQI
51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISYMYLA PVFGYLGDRY
101 NRKYLKCGGI AFWSLVTLGS SFIPGHEFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRLSIFY FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LFLVVRPEP RGAVERHSDI PPLNPTSWWA DLRLARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRA DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCATFVGALG
401 GAAFLGTAFI IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1e15, frame 1

TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,
 N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid
 C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,
 N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11_18 gene: "F6H11.180"; product: "predicted protein";
 Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII
 project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N
 = 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9
 Length = 488

HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISYMYLAPVFGYLGDRYNRKYLMCGGIAFWSLVT 117
 + ++ V Y N + + + VF+ S+MV +PV GYLGD+NRK+M G+ W
 Sbjct: 29 AGVLTQVQTYYNISDSLGLLIQTVFLISEFMVSPVCGYLGDRFNKRWIMIGVGIWLGAV 88

Query: 118 LGSSFIPEGHEFWLLLLTRGLVGVGEASYSTIAPTILADLFVADQRSRLSIFYFAIPVGS 177
 LGSSF+P HFWL L+ R VG+GEASYS +AP+LI+D+F +RS + IFYFAIPVGS
 Sbjct: 89 LGSSFVPANHFVFLVLRSFVGIGEASYSNVAPSLISDMFNGQKRSTVFIMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRPEPRGAVER---HSDLPL 233
 GLG+I GS V + G W W +RV+ G++ ++ L L EP RGA ++ D+
 Sbjct: 149 GLGFIVGSNVATLTGHVQWQIRVSAIAGLIVMIALVLFTYEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWADLRALARNLIFGLITCLTG 259
 T++ DL L + L+ C G
 Sbjct: 209 TNTTYLEDLVILLKTPT--LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGVGLGVEISRRL-----RHSNPRADPLVCATGLLGSAPFLFLSL 300
 L FG IT G+GV G +S+ L R RA PLV G L +APFL + +
 Sbjct: 277 LYFGAITTAGGLIGVIFGSMLSKWLVAGWGPFRRLQTDRAQPLVAGGALLAAPFLIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

Pedant information for DKFZphmcf1_1e15, frame 1

```
[LENGTH]      454
[MW]           49013.35
[pI]           7.66
[HOMOL]        TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-51

[BLOCKS]       BL01022D
[PROSITE]      MYRISTYL             11
[PROSITE]      CAMP_PHOSPHO_SITE     1
[PROSITE]      CK2_PHOSPHO_SITE      3
[PROSITE]      PROKAR_LIPOPROTEIN    1
[PROSITE]      GLYCOSAMINOGLYCAN     1
[PROSITE]      PKC_PHOSPHO_SITE      4
[KW]           TRANSMEMBRANE 8
[KW]           LOW COMPLEXITY        15.42 %
```

556

SEG
PRD hhhhhhhcccccccccccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMM.....

Prosite for DKFZphmcf1_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1e15.1)

DKFZphmcfl_lgl3

group: mammary carcinoma derived

DKFZphmcfl_lgl3 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKFZ

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```

1  GAAACCTGAT CTCATAAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51  GACCCCTATTT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACTACA GATTGTCTC CAGGGGCTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATTT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
251 CTACAACTAT GAATGAGAGA GCCTTATTGT CATCGTATTT AGTTGCAATAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA CGGCTGAAAT AAATTATCCT
351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATTT GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCCACTCT CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTTAACT
701 GGAACATTG TAAAGGAATT TCAAGTCATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACATGCG
801 TGTTTGGAAT CACTGTTTGA TTCATCGAGA AGCTTTGGTA TCCAAAGAAA
851 TTTACCAAG TCTGATGGAT GTATTGAAA ATGCAGTGAA AACTGTTAAT
901 TTTATTAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTTGTTT
951 AGAGATTGGA GTGAACACA CCCACTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAC CAGGAACGAG
1051 ATTTACATT TTCTCGTTGA AAAGCAATCT CATTGGCAA ATATTTTGTA
1101 AGACGACATT TGGGTAAACA AATTGGCATA TTTAAGTGAT ATTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAAAAACAA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGTCTCAAA CTTTTAATTA
1401 TACTTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAA CAATAATTGA GTTAAACTTG
1501 GAGCCTGAAG AAGAGAAATG ATTATTGCAG CTCAGTTTAT CATTACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACAT CATATTGTG TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAAAG AAGAGAAATA GGCTCAATAG TGCACAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTCTT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCATCAC ATTAATATCA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAAC CCCATCTCTA
1951 CTAATAATAG AAACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCTGAGGC AGGAGAAATCT CTTAAACCAAG GAAAGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGTGCA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAAATTGT ATTTGTACTT TTAAGGGAT
2151 TTTGAGTAT GTTGTAGTTA AACGTTAATA AAATTATATT TGTAATTAGG
2201 AAAAAA
```

BLAST Results

Entry AC005020 from database EMBL:
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.
Score = 9110, P = 0.0e+00, identities = 1822/1822

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573
Category: similarity to unknown protein

1 MTPESRDTTD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
51 SRSTTMNERA LLSSYLVAIR VAKEKMAHTA AEKILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKKHKGIS SDGTANMTGK HSRLTEKLEL ATHNNVWNN CFHREALVS
251 KEISPSLMDV LKNAVKTVMF IKGSSLNSRL LEIFCSEIGV NHTHLLFHTF
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFQK TLLWQARLK SNRPSYMF
401 TLLQHIENI INEDCLKEIK LEILLHLTSL SOTFNYYFPE EKFSLEKNI
451 WMKDPFAFQN PESIIELENE PEEENELLQL SSSFTLNYY KILSLAFWI
501 KIKDDPPLLS RKSILLPLF TTYLCELGF SILRLKTKK RNRLNSAPDM
551 RVALSSCVDP WKELMNRQAH PSH

BLASTP hits

Entry AC004877_3 from database TREMBLNEW:
gene: "WUGSC:H DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens
PAC clone DJ0751H13 from 7q35-qter, complete sequence.
Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211_1 from database TREMBL:
product: "Hermes transposase"; Musca domestica Hermes transposase
gene, complete cds.
Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

Alert BLASTP hits for DKFZphmcf1_lg13, frame 1

TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P
= 1.1e-23

>TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
sapiens mRNA for KIAA0766 protein, complete cds.
Length = 607

HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23
Identities = 120/485 (24%), Positives = 229/485 (47%)

Query: 89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+
Sbjct: 124 CMEVLLREVLPEH-VSVLQGVLDSPDITRQILSIDRNLRLNQLFNARDFKAYSLALDDQ 182
Query: 148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
+A LLV++R V + + EDLL +NL H + G + LE+ L L+ +
Sbjct: 183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHFVSGALMSAILES--LQTAGLSLQR 240
Query: 206 KGISSDGTANMTGKHSRLTEKLEATHNNVWNN--HC--FIHREALVSKEISPSLMDVL 261
G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct: 241 MVGLTTHTLRMIGENSGLYSYMREKAVSPNCWNVIHYSGLHLELLSSYDQVDN--QII 298
Query: 262 KNAVKTVMFVIGSSLSNRLLEIFCSEIGVNHHTLHFTFTEVR-WLSQGVLSRVYELRNEI 320
+ + IK + + +E H + + WL +GK L ++ LR E+
Sbjct: 299 NTISEWIVLIKTRGVRRPEFQTLTSESEHGERVNGRCNLNWLRRGKTLKLFSLRKEM 358
Query: 321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFGILNELSLKMQGKNNDIFQYLEHILGFQK 380
FLV + + + F D W+ +L DI L ELS +++ +HI F+
Sbjct: 359 EAFLVSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417

Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22
Identities = 120/485 (24%), Positives = 228/485 (47%)

Pedant information for DKFZphmcf1_lg13, frame 1

560

[illegible]

Prosite for DKFZphmcf1_lg13.1

PS000001	216->220	ASN GLYCOSYLATION	PDCC000001
PS000001	291->295	ASN GLYCOSYLATION	PDCC000001
PS000005	116->119	PKC PHOSPHO_SITE	PDCC000005
PS000005	218->221	PKC PHOSPHO_SITE	PDCC000005
PS000005	225->228	PKC PHOSPHO_SITE	PDCC000005
PS000005	358->361	PKC PHOSPHO_SITE	PDCC000005
PS000005	391->394	PKC PHOSPHO_SITE	PDCC000005
PS000005	445->448	PKC PHOSPHO_SITE	PDCC000005
PS000005	485->488	PKC PHOSPHO_SITE	PDCC000005
PS000005	510->513	PKC PHOSPHO_SITE	PDCC000005
PS000005	538->541	PKC PHOSPHO_SITE	PDCC000005
PS000006	55->59	CK2 PHOSPHO_SITE	PDCC000006
PS000006	79->83	CK2 PHOSPHO_SITE	PDCC000006
PS000006	95->99	CK2 PHOSPHO_SITE	PDCC000006
PS000006	136->140	CK2 PHOSPHO_SITE	PDCC000006
PS000006	183->187	CK2 PHOSPHO_SITE	PDCC000006
PS000006	189->193	CK2 PHOSPHO_SITE	PDCC000006
PS000006	256->260	CK2 PHOSPHO_SITE	PDCC000006
PS000006	445->449	CK2 PHOSPHO_SITE	PDCC000006
PS000006	463->467	CK2 PHOSPHO_SITE	PDCC000006
PS000006	546->550	CK2 PHOSPHO_SITE	PDCC000006
PS000007	364->372	TYR PHOSPHO_SITE	PDCC000007
PS000008	137->143	MYRISTYL	PDCC000008
PS000008	273->279	MYRISTYL	PDCC000008
PS000008	289->295	MYRISTYL	PDCC000008

(No Pfam data available for DKFZphmcf1_1g13.1)

DKFZphtes3_14g5

group: testes derived

DKFZphtes3_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and and nuclear localization signals of lyar.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```
1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTACCTA AAAATGGTAT
151 TTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGAG AAACGTGTGA TGCTTTTCT GCATTGACTG
251 CGGTAAAGAT TTCTGGGGCG ATGACTATAA AAACACGCTG AATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAGGCT ATGAAGGTAA AACCCACAAA
351 GGCACATCA AACAGCAGGC GTGGATTGAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCA AAGTGAGAGA ACTTTAGAG CAAATTAGTG
451 CTTTTCAGAA CGTTCCAGG AAAAGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTTCAGAA GCTTCCAAAC GCGAACCATG CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAGA GAAAGAAAGG AAGAACGCA GAAGAAAGG AAAAGAGAAA
751 AGAAGAAGT AAAATTAGAA AACCACAGG AAACTCAAG GAATCAGAAG
801 CCTAAGAAGC GCAAAAAGG ACAGGAGGCT GACCTGAGG CTGGTGGGGA
851 GGAAGTCCCT GAGGCCAATG GCTCTGAGG GAAGAGGAGC AAGAAGAAGA
901 AGCAGCGCAA GGACAGGCC AGTGAGGAG AGGCAGCGT GGGCGCAGGG
951 AAGAGCAAGC GGAGGCACCT GGAAGTTGAA ACAGATTCTA GAAGAAGAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGGCGG AGAACAGAA GACGATGAGG
1051 CTCCTGCAAA AGGTAAATTC AACTGGAAGG GAATATTAA AGCAATCTCT
1101 AAACAGGCCC CAGACATGTA AATAACCATC AAAAAGCTAA GGAAGAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCACAGA TCCGAAGAGG
1201 AACTCCTGGT CATCTTTAAC AAGAAAATCA GCAAGAACCC TACCTTTAAG
1251 TTATTAAAGG ACAAGTCAA GCTTGTGAAA TGAACATTG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTCTTCTCT TCACTGCTGT TTATAAATG
1351 TGTAATGAAT TCTAACAAC CAAATTTTC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAATAT
1451 ATCTGGTCC AAACCTCAAA AAAAAAAGG AAAAAAAGG AAAAAAAGG
1501 AAA
```

BLAST Results

No BLAST result

Medline entries

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379
Category: strong similarity to known protein
Classification: Cell division
Prosites motifs: ATP_GTP_A (60-68)

1 MVFFTCNACG ESKKIQVEK HVSVCNCEC LSCIDCGKDF WGGDYKNHVK
51 CISEDQKYG KGYEGKTHKG DIKQAWIQ ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNMKN SLKVHNSIL DQVWNIFSEA SNSEPVNKEQ
151 DORPLHPVAN PHAEISTKVP ASKVKDAVEQ QGEVKNKRE RKEERQKRRK
201 REKKELKLEN HQENSRNQPK KRRKKQGEAD LEAGGEEVPE ANGSAKRRSK
251 KKKQRKDSAS EEARVAGAK KRRHSEVET DSKKKMKLP EHPEGGEPEP
301 DEAPAKGKFN WKGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHRS
351 EEEELLVIFNK KISKNPTFKL LKDKVKLVK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058_18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse
Length = 388

HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144
Identities = 275/388 (70%), Positives = 317/388 (81%)

Query: 1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYG 60
MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGGDYK+HVKCISE QKYGG
Sbjct: 1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGGDYKSHVKKCISEGQKYGG 60
Query: 61 KGYEGKTHGDIKQAWIQI+ELIK+PNVSPKVRELL+QISAFDNVPRKAKFQNMKN 120
KGYE KTHGDI KQAWIQI+ELIK+PNVSPKVRELL+QISAFDNVPRKAKFQNMKN
Sbjct: 61 KGYEAKTHGDAKQAWIQINELIKPNVSPKVRELLQISAFDNVPIKAKFQNMKN 120
Query: 121 SLKVHNSILDQVWNIFSEASNSPVNKEQDORPLHPVANPHAEIS-TKVPASKVKDAVE 179
SLKVH++S+L+QVW+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct: 121 SLKVHSDSVLEQVWDFSEASSSE---QDQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176
Query: 180 QQGEVKNKRRERKEERQKRRKREKKELKLENHQENSRNQPKRRKKQGEADLEAGGEEVP 239
+Q E KKNKRRERKEERQK RK+EKKELKLENHQEN R QPKRRKK QEA EA GE+
Sbjct: 177 EQTEAKKNRRERKEERQKRRKREKKELKLENHQENLRGQKPKRRKKQGEAGHEAAGEDGA 236
Query: 240 EANG-----SAGRSKKKKQRKDSASEE-----RVGAGKRRK-RHSEVETDSKKKKM 287
+ +G G+ S++ R E+ A + AGKRRK +HS E+ KKKKM
Sbjct: 237 DSGPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRRKPKHSGAESGYKKKKM 296
Query: 288 KLPEHPEGGEPEDEAPAKGKFNWKGTIKAILKQAPDNEITIKLRKKVLAQYTVTDEH 347
KLPE PE GE +D EAP+KGKFNWKGTIKAILKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct: 297 KLPEQPEEGEAKDHEAPSKGKFNWKGTIKAVLRQAPDNEISVKKLKKKVIAQYHAVMNDT 356
Query: 348 HRSEELLVIFNKKISKNPTFKLLKDKVKLVK 379
EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct: 357 SHHEELLVIFNKKISKNPTFKLVKDRVKLLK 388

Pedant information for DKF2phtes3_14g5, frame 3

Report for DKFZphtes3_14g5.3

[LENGTH] 379
[MW] 43634.03
[pI] 9.59
[HOMOL] PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
[BLOCKS] BL00603D Thymidine kinase cellular-type proteins
[BLOCKS] BL00530C
[PROSITE] ATP_GTP_A 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 18.73 %

SEQ MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGDYKNHVKCISEDQKYGG
SEG
PRD cccccccccccccchhhhhhhheeeccccceccccccccccccceeecccccccc
SEQ KGYEGKTHGDIKQAWIQIKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN
SEG
PRD cccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhcccccchhhhhhhhhhhc
SEQ SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEISTKVPASKVKDAVEQ
SEG
PRD cccccchhhhhhhhhhhhhhhccchhhhhhhhhccccccccccccceccccchhhhh
SEQ QGEVKKNKREERQKKRKREKELKLENHQENS RNQPKKKKKGQADLEAGGEEVPE
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhchhhhhcccccc
SEQ ANGSAKRSKKKQKQKDSASEEARVAGAKRRRHSEVETDSKKKKMLPEHPEGGEPEP
SEG
PRD cccccchhhhhhhhhccchhhhhhhhhccccccccccccccccchhhhhcccccccccc
SEQ DEAPAKGKFNWGTIKAILKQAPDNEITIKLRKKVLAQYTYTDEHHRSEELLVIFNK
SEG
PRD cccccceehhhhhhhhhccccccccchhhhhhhhhhhhhccchhhhhhhhhhhhh
SEQ KISKNPFTKLLKDKVKLVK
SEG
PRD cccccchhhhhhhhhcc

Prosite for DKFZphtes3_14g5.3

PS00017 60->68 ATP_GTP_A PDOC00017

(No Pfam data available for DKFZphtes3_14g5.3)

DKFZphtes3_14h21

group: nucleic acid management

DKFZphtes3_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNATg

Sequenced by BMF2

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGAA CAATGTCCCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GTCGTTGCT AGTCGCGGAA GCTCGACAGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGCGGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGGATA TAGTGTCCGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
201 CCCCAGGAGG CCGTGGCCGC TGGTCACGAG GAACGCGCGC TGTGTTTTCG
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTCTGGGT GGGTCAAAAA
301 TAAAGAAATAT ACAAGTACA ACAACACCA CAATCCAAAT AATACAAGAA
351 CAACCAGAAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCACACGAA
401 AGCAAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAATATACA
451 ATTCAGAAATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
501 GGAAGCAGCAG ATAACAATGT TGTGACAGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAAT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTTATA AAGAGTCCAC TGCCACAAAT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTTAATAT
701 AACGTGGGAT GACTTGAAGG ATGGGGAGAA ACGACCTATC CCCAATCCTA
751 CTTGCACATT TGATGACGCC TTCAATGTT ATCCTGAGGT TATGGAAAAC
801 ATTAATAAGG CAGGTTTCA AAAGCCAACA CCTATTCACT CACAGGCATG
851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATTGTGTTAT TTAATGCCCT GATTATTCA TCTGGTCTCT
951 CAACCCAGCC TTAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAAAATAT
1051 CATATAAAGG GCTTCGGACT GTTGTGTAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTCTGCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTTAGAT GTGCGCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCTTCATTC AGTTTCATCGC CTCGCACAAT
1351 CTTATTGTA AGAACCAATG ATTGTCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAGTT CAGTGAAGCA AATATAATT GTAACCAACG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAGG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGTCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAACAGGCG AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGGTCTT TCCATTACAA
1801 CTTTGACTAG AAATGATTGG AGGGTTCCTT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAAT GGAAGAGAAA ATGGAAAGAC
1951 CTCAGGAGAG GCCCAAGAGG TTTTATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAAATCA AGATTTTATA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAATAATAA
2101 GTGTTTGAAA ATATAGAAATC CAGTGTTTAA TACTTTCTTT AATAAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648
 Category: strong similarity to known protein
 Classification: Nucleic acid management
 Prosite motifs: ATP GTP A (286-294)
 DEAD_ATP_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTTNTT
101 IQIIQEQPES LVKIFGSKAM QTKAKAVION FVKKLEENYN SECCIDTAFO
151 PSVGKDGSTD NNVVAGDRPL IDWDQIREEG LKWKQTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPINPT CTFFDAFCQY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTGT GKTLCYLMFG
301 FIHLVLQPSL KGQRNRPGL VLTPTRELAL QVEGECKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDIII ATPGRLNDLQ MSNFVNLKNI TYLVLEADK
401 MLDMGFEFQI MKILLDVRPD RQTVMTSATW PHSVHRLAQ VLEPHIVYV
451 GTLDLVAVSS VRQNIIVTTE EEKWSHMOTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQDRREKAL ENFKTGKURI LIATDLASRG
551 LDVHDVTHVY NFDPRNIEE YVHRIGRTGR AGRTGVSIIT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRFKKPH
    
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14h21, frame 3

TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP887_16 gene: "dbp2"; "SPBP887.16c"; product: "p68-like
 protein."; S.pombe chromosome II pl p887., N = 1, Score = 971, P =
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (Saccharomyces cerevisiae), N = 1,
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces
 pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid
 Y54G11A
 Length = 504

HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREGLKWKQTKWADLPPIKKNFYKSTATSAMSKEADSWRKENFNITWDDLKDGEK 233
      D++++E W K P I ++ YK +S + + ++
Sbjct: 23 DRLKDFNSWMK-----PIVRDLYKIPNEQKNLSPEQLQELTYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFFDAFCQYPEVMENIKKAGFQKPTPIQSQAQWPIVLQGLDQIGVAQTGTGKT 293
      IP P +F+ AF +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASINGEIRKNGFEKPSPIQSQMWPLLSGGQDCIGVSGTSGGNT 135

Query: 294 LCYLMFGFIHLVLQPSL----KGQRNRPGLVLTPTRELALQVEGECKYSYKGLRSVC 348
      L +L+P +H+ Q + + Q+ P +LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALLHIDAQLAQYEKNDEEQKPSFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGNRDEQIEELKKGVDIIATPGRLNDLQMSNFVNLKNITLVLEADKMLDMGFEP 408
    
```



```

      +YGGG+R EQ+E  + GV+I+IATPGRL DL      ++L ++TY+VLDEAD+MLDMGFE
Sbjct: 196 LYGGGSRPEQVEACRGGVEIATPGRLTDLSDNGVISLASVTYVVLDEADRMLDMGFEV 255

Query: 409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLEKPMIVYVGTLDLVAVSSVKQNIIVT 468
      I ++IL ++RPDR  +TSATWP V +L  Y KE ++  G+LDL +  SV Q
Sbjct: 256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKAEMAVNGSLDLTSCSKSVTQFFEFV 315

Query: 469 TEEKW---SHMQTFLOQMSSTD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
      + ++      + FL +      + K+I+V K +ADHLSSD + I+ + LHG R Q
Sbjct: 316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSKVMADHLSSDFCMKGINSQGLHGGRSQS 375

Query: 525 DREKALENFKTKGVRIIATDLASRGLDVHDVTHVYVDFPRNIEEYVHRIGRTGRAGRT 584
      DRE +L  ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct: 376 DREMSLNMLRSGEVQILVATDLASRGIDVDPDITHVLNDFPMDIEEYVHRVGRGTGRAGR 435

Query: 585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMARFEAHQRKREMERKMERPOGRF 644
      G +++ L ND  LI ILE++ Q +P++L AE++      K + R  RP R
Sbjct: 436 GEAMSLFWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KCQSGRDGPRPSFRN 492

Query: 645 KK 646
      K
Sbjct: 493 NK 494

```

Pedant information for DKFZphtes3_14h21, frame 3

Report for DKFZphtes3_14h21.3

```

[LENGTH]      648
[MW]           72873.51
[pI]           8.84
[HOMOL]        TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-
101
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-70
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
[FUNCAT]       1 genome replication, transcription, recombination and repair [H.
influenzae, H10892] 2e-49
[FUNCAT]       j mrna translation and ribosome biogenesis [H. influenzae, H10231 RNA] 1e-48
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 1e-07
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 4e-96
[PIRKW]        RNA binding 3e-87
[PIRKW]        DEAD box 5e-50
[PIRKW]        transmembrane protein 4e-27
[PIRKW]        DNA binding 3e-67
[PIRKW]        recF recombination pathway 3e-10
[PIRKW]        ATP 4e-96
[PIRKW]        purine nucleotide binding 5e-50
[PIRKW]        P-loop 4e-96
[PIRKW]        hydrolase 9e-45
[PIRKW]        protein biosynthesis 5e-50
[PIRKW]        ATP binding 1e-61
[SUPFAM]       WW repeat homology 8e-88
[SUPFAM]       DEAD/H box helicase homology 4e-96
[SUPFAM]       unassigned DEAD/H box helicases 7e-87
[SUPFAM]       ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM]       ATP-dependent RNA helicase DHH1 2e-43
[SUPFAM]       recQ protein 3e-10
[SUPFAM]       Bloom's syndrome helicase 5e-07
[SUPFAM]       translation initiation factor eIF-4A 5e-50
[SUPFAM]       recQ helicase homology 3e-10
[SUPFAM]       tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE]      DEAD_ATP_HELICASE 1

```

[PROSITE] ATP_GTP_A 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] KH domain family of RNA binding proteins
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 8.49 %

```

SEQ  MSHHGGAPKASTWVVASRRSSTVSRAPERRPAEELNRTGPEGYSVGRGGRWRTSRPPEA
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ  VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNIQSTNTTTIQLIQEPESLVKIFGSKAM
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  CCCCCCCCCchhhhhccceeeccccccccccccccccccccccccccccccccccccchh

SEQ  QTKAKAVIDNFVKLEENYNSECGIDTAFQPSVGKDGSTDNVNVAGDRPLIDWDQIREEG
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ  LKWQKTWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGKRPINPT
SEG  .....
PRD  chhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhheeecccccccccccccc

SEQ  CTFDADFQCYPEVMENIKKAGFQKPTPIQSQAQWPIVLOGIDLIGVAQTGTGKTLCLYLMPG
SEG  .....
PRD  CCCCCCCCCchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ  FIHLVLQPSLKGQRNRPGLVLTPTRELALQVEGECCKYSYKGLASVCVYGGGNRDEQIE
SEG  .....
PRD  eeeeeccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  ELKKGVDI IATPGRNLDLQMSNFNVLKNITYLVLEADKMLDMGFEPQIMKILLDVRPD
SEG  .....
PRD  hhhheeeccccccccchhhhhhhccccccccccccccccccccccccccccchhhhhhhhhcc

SEQ  RQVTMSATWPHSVHRLAQSYLKEPMIVYVGTLDLAVSSVKQNIIVTTEEEKWSHMOTF
SEG  .....
PRD  ceeeeccccchhhhhhhhhhhhhheeeccccccccccccccccccccccccchhhhhhhhh

SEQ  LQMSSTDKIVIVFSRKAVADHLSSDLILGNISVESLHGDRQDRKALENFKTGKVRI
SEG  .....
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LIATDLASRGLDVHDTVHYNFDFFRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS
SEG  .....XXXXXXXXXXXXX.....
PRD  eeehhhhhhccccccccccccccccccccccccccccccccccccccccccccccccchhh

SEQ  ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRPKKFH
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc

```

Prosite for DKFZphtes3_14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphtes3_14h21.3

HMM_NAME	DEAD and DEAH box helicases	
HMM	*gLpPWILRnIyeMGFEkPTPIQQaIPiLeGRDVMACAQTGSgKTAAF	
Query	248	QCYPEVMENIKKAGFQKPTPIQSQAQWPIVLOGIDLIGVAQTGTGKTLCLY 296
HMM	L+P ++H+ +P +++ Q+ P +L+L+PTRELA+Q++ EC K+++ +	
Query	297	LMPGFIHLVLQP-SLKGQRNRPGLVLTPTRELALQVEGECCKYSYK-G- 343
HMM	IRImCIYGGtnMRdQMRnLeRGpPHIVATPGRLLDHIERgtldLDRIeM	
Query	344	LRSVVCYGGGNRDEQIEELKKGV-DIIATPGRNLDLQMSNFNVLKNITY 392
HMM	LVMDEADRLMDMGFIDQIRrIMrQIPmwnRQTMFSATMPDeIqELARr	
Query	393	LVLDEADKMLDMGFEPQIMKILLDVR--PDRQVTMSATWPHSVHRLAQS 440

```

HMM          FMRNPiRiNId.MdElTtnEnIkQwYiyVerEMWkfcdLcrLIe*
          ++++P  + ++  D  +++  +KQ +I+  E++K +  ++++
Query      441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-BEEKWSHMQTFLQ  482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
          + + +++++G++IG+GGS I++I++ ++++I I++E+ +  + + I
Query      71  CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQE-Q-P---ESLVKIF  115

HMM          G*
          G
Query      116 G    116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWLknl....GlrvmYIHGdMpQeERdeIMddFNnGEynVLicTD
          + +++ L+ +  +I+V ++HGD++Q+R+++++F++G+ ++LI+TD
Query      497 KAVADHLSSDLILGNISVESLHGDRDREKALENFKTGKVRILIATD  545

HMM          VggRGIDIPdVNVHVINYDMPWNPEqYIQRIgRTgRIG*
          +++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query      546 LASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAG  582

```

DKF2phtes3_14p14

group: testes derived

DKF2phtes3_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMF2

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```
1 GAAGCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTC
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTG
101 GGTGCTGAAG GCCAGGCAGA GCATTGGCT GTAGGAGGGC CGATCCTCCT
151 CGGGCCTGTT ACCGGCGGCT CTTTGTCTT AGACCTGGGG TTCTGGGCTT
201 CACGGATTCC AAGGAATGGA ACCTTGGGCC ATGCGTGTGA ACGAGCTCTA
251 TGTGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGCTCTA
301 ACATCAGTTT ACCCAATCTG CACTGCGAGT TGGTTGGGCT TGACATTGAG
351 GATGAGATGG GCAGGCACGA AGTGGGCCAC ATCGACAAC CCATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGGCCGTT CGAGGGGGCAG TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTCTCT
501 GCTGTCTCCC CGCTCCCTGG AAACCTGGTTG TGGAGGCAC CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TTCTGTGCCA GCTGCCTGGC
601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTTCAG CCCCCAGCCC TGGGTTCAAG
751 TCCAGCTCT ACCCTTCTTT GGGCCCTACA AGTCACCTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
851 AGTTTGTGTG GGGATTAAAT GAGCTAATGC AGATTTCATT ATTGAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CAGCGTTCCT GGGCTGGTGG GGCTCCCAT TCGGTAGAGG GAGCAGCTCT
1001 ACAAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGGT GGGCTTGAAA GTACCTTGTC CGCTCAGAAG
1101 GACCAATCAA GGTTCAGTGT TGTTTGTCC TCAGAACCAAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACGCTGTGAG GCAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGGCCAGAGA GGTAAAGTGT CTTGCCCGAG
1351 GTCACACAGC TATCTGCATG TCCCACAAC CTCTTCCCA GCGCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCCCT CAGACCTGA CACCTGAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTGTTTG TCCAGTATAG
1551 CATTTAACTG CATCAGAAC AGTATGAAA GACCAGGAAT CCAGATTCTT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCGCT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAG TGGGCAGTGT TCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCCTG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGCTAAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGTCT
1851 GGGAGAAATC TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACACAGCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGATGCT GGGGGTCTG
1951 GTTCACTGGG GGCTCTCTCT GCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCCGTA ATTTGGATTT TATGGTGGTT GTGATGAAA GCCATTTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGAATCAGA TTTATGCCCT AGAATGGAC
2101 TATCCAATAG GTTGCCAGCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTAC AATTCACTCA CTCATTCTCA TCAACCACAT TTCAAGTGCT
2201 CAAAGCCAGC TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCAGGAG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCTCTACT AAAAAATCAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCGTAC CACTGCATCT CAGTCTGGGC GACAGAGTGA GACTCCATCT
2601 CAAACAAAC AAAAAGGA TGGGGCTGG CTGGAGAGGG TGCGAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGTAAGGG CACATGTGGA
```

```

2751 AGCAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCTGTGT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACTG GTCATTCCAC TCTGGTTTCC TTGGGGTTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAAATTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTTCTGG ACCACAGATG TTGAGTTGGG AGTCATTAGT GAATAGATGA
3001 TGTGGAAAGT CATGGGTCTT CAGAGTGGGG GTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTGTGCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG
3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCCAGCAC TTTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCCTG GTCAACATGT
3601 CAAACCCCTG TATCTACAAA AAAATACAAA AGTTAACCAA GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGCTT CTGAGCCAGC
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAACCCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTAA AATATTTTAA TTACTATTTA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159
 Category: putative protein
 Classification: no clue

```

1 MERWAMRVNE LYVDDPKDSD GGGKIDVSLNI SLPNLHCELV GLDIQDEMGR
51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPCLSPFY LLPFAVSP
101 PGNWLWRHSL DLTLTQPPAS EGSCPAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSLQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p14, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3_14p14, frame 3

Report for DKFZphtes3_14p14.3

```

(LENGTH) 159
(MW) 17778.55
(pI) 5.74
(FUNCAT) 99 unclassified proteins [S. cerevisiae, YAL042w] 5e-04
(KW) Alpha_Beta

```

```

SEQ MERWAMRVNELYVDDPKDSDGGKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM
PRD ccchhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
SEQ KIPLNNGAGCRFEGQFSINKVWKPCLSPFYLLPFAVSPPLGNWLWRHSLDLTLTQPPAS
PRD eeeeeeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc

```

WO 01/12659

PCT/IB00/01496

SEQ EGSCPAAWPFLLRIWMCVQAPWGFKPLMAGSGRSYSSLQ
PRD cccccchhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3_14p14.3)

(No Pfam data available for DKFZphtes3_14p14.3)

DKFZphtes3_l4p7

group: testes derived

DKFZphtes3_l4p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1  GGAATCCAAA  GAAACAGTTA  TGATGGGGGA  CTCTATGGTG  AAAATAAATG
51  GGATTATTTT  AACAAAATCA  AATGCTATTT  GCCACTTAAA  GAGTCACCCA
101  CTTCAGCTAA  CTGATGATGG  AGGCTTCAGT  GAAATAAAGG  AGCAAGAAAT
151  GTTCAAAGGA  ACAACATCTT  TACCATCTCA  TCTCARGAAT  GGAGGGGACC
201  AGGGGAAGAG  ACATGCGAGG  GCCTCATCAT  GCCCCAGTAG  CTCAGACCTG
251  AGCAGGCTGC  AAACCAAAGC  AGTCCCAAAA  GCTGACCTGC  AAGAAGAGGA
301  CGCAGAAATA  GAAGTAGACG  AAGTCTTTTG  GAATACAAAG  ATTGTACCGA
351  TTTTGCCTGA  ATTAGAAAAG  GAAGAAAACA  TTGAAACGGT  TTGTGCTGCT
401  TGCACACAA  TTATCATATG  TTTAGAGGAA  GGAACATGTC  TTGGAATAAA
451  ATTTAAGGGA  AGAAGTATTC  TCCTGAAGAC  CCTGTGTAAA  CTAGTTGATG
501  TTGGTTCAGA  CTCGCTCAGC  CTTAAACTTG  CAAAAATAAT  TCTAGCACTT
551  AAAGTGAGTA  GAAAGAATCT  TCTTAATGTC  TGCAAACTTA  TATTTAAAT
601  TAGCAGGAAT  GAGAGAATG  ATTCTTTGAT  TCAAAATGAC  AGCATTCTGG
651  AATCATTATT  GGAGGTACTA  AGAAGTGAAG  ACCTGCAAA  TAACATGGAA
701  GCTTTTTTAT  ACTGTATGGG  GTCTATAAAG  TTCATTCTG  GAAATCTGGG
751  ATTTCTTAAT  GAAATGATCA  GCAAAGGTGC  TGTGGAATA  CTGATAAATT
801  TGATAAAACA  AATAAATGAG  AACATCAAGA  AATGGGTAC  ATTTTGGCT
851  AATTCTGGCC  ACTTGCTAGT  CCAGGTGACT  GCTACATTGA  GAAACTTGGT
901  TGATTCTACA  TTAGTAAGAA  GTAAGTTCCT  AAACATCAGT  GCCCTTCCCC
951  AGCTCTGCAC  GGCAATGGAA  CAGTACAAGG  GTGACAAGGA  CGTCTGTACC
1001  AATATTGGCA  GAATATTCAG  CAAACTTACT  TCTTACCGTG  ACTGCTGCAC
1051  AGCCTTGGCC  AGCTATTCCA  GATGTTATGC  CTTATTCTG  AATCTAATTA
1101  ACAAAATCCA  GAAGAAGCAG  GATTTAGTCG  TCCCTGTTGT  TTTTATTTCT
1151  GGCAACCTGA  CGGCAAAAAA  TAACCAGGCT  CGTGAACAA  TTTCCAAAGA
1201  GAAAGGGAGC  ATCCAAACTC  TGCTGTCAAT  ATTCCAGACG  TTCCATCAGC
1251  TGGATCTGCA  TTCCAGAAG  CCGTGGGCC  AACGAGGCGA  GCAGCACAGG
1301  GCGCAGAGGC  CGCCGTCAGA  GGCAGAGGAC  GTGCTCATCA  AGCTGACTCG
1351  TGTGCTGGCC  AACATTGCCA  TCCACCCGGG  CGTGGGCCCG  GTGCTGGCCG
1401  CCAACCCGGG  GATAGTGGC  CTGCTCCTGA  CCACGCTGGA  ATACAAGTCA
1451  CTTGATGATT  GTGAGGAGCT  GGTGATCAAT  GCTACAGCGA  CAATCAACAA
1501  TTTATCTTAC  TACCAAGTGA  AGAATTCAT  AATTCAAGAC  AAAAAGCTAT
1551  ATATTGCTGA  ATTGCTCTTA  AAGCTTCTTG  TCAGTAACAA  CATGGATGGA
1601  ATCCTGGAGG  CTGTGCGTGT  TTTGGGAAAT  CTCTCCAGG  ACCATGATGT
1651  CTGCGATTTC  ATTGTGCAGA  ACAATGTCCA  CAGGTTTCATG  ATGGCGCTGC
1701  TGGATGCTCA  GCATCAGGAT  ATCTGCTTTT  CTGCTGTGG  TGTCTCTCTC
1751  AATCTCACTG  TGGATAAAGA  CAAGCSTGTC  ATCTTGAAAG  AAGGAGGTGG
1801  CATTAAAAAG  TTAGTGGAGT  GTTAAAGAGA  TTTGGGTCT  ACTGATTGGC
1851  AGCTGGCCTG  CTTGGTTTGT  AAAACTTTAT  GGAACCTCAG  TGAAAACATC
1901  ACTAATGCTT  CGTCATGTTT  TGGAAATGAA  GACACCAACA  CACTCTTACT
1951  CTTGCTCTCA  TCATTTTAT  ATGAAGAACT  AGCACTGGAT  GGCAGTTTGT
2001  ATCCAGACCT  AAAAACTAT  CACAAACTCC  ATTGGGAAAC  AGAATTCAAA
2051  CCGTGGGCAC  AGCAGCTTCT  AAACCGAATT  CAGAGACATC  ACACCTTCTC
2101  GGAACCCCTG  CCCATTCCT  CTTCTAACA  TGATGCAGAT  TAACAGTAGA
2151  AACGAGAACT  CACGTCTCCC  TCATTCTTAA  GAACGTGTAA  CAACGTGAA
2201  CATTTTTTTC  AGCATTAACA  AATGTGGAAA  GTTTTCAAG  AACTGGTTTT
2251  AGTGAGTAGC  TGAAGTATTT  TTTAAATTA  AGCATTCTCT  CTTGTTAGGT
2301  ATTATGAAA  AATGAATATA  CACATTATAT  TTCCTGTGA  GAGAAATGTA
2351  AGATCAAAAT  ATGTGCATTT  TCAAGTAAAT  GACTTTTCT  TCTATTCCTC
2401  ATTAACAAAT  TTAGTCTAG  TCTTAAAAAA  AAAAAAATA  AAAAAAATA
2451  AAAAAAATA  AAAAAAATA  AAAAAAATA  AAAAAAATA  AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702
Category: putative protein

```
1 MMGDSMVKIN GIYLTGSNAI CHLKSHPQLQ TDDGGFSEIK EQEMFKGTTS
51 LPSHLKNGGD QGKRHRARASS CPSSDLSRL QTKAVPKADL QEEDAIEVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHALEEGNM LGNKFGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IFKISRNEKN
201 DSLIQNDSIL ESLLLEVLRSB DLQTNMEAFI YCMGSIKFI SGNLFLNEMI
251 SKGAVEILIN LIKQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSLSVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RYALFLNLI NKYQKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQI
401 LLSLFQTFHQ DLHLSQKPVG QRGEQRAQR PPSEADVLI KLTRVLANTA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSSYYQV
501 KNSIIQDKKL YIAELLKLL VSNMMDGILE AVRVFVGNLSQ DHDVCDFIQV
551 NNVHRFMAL LDAQHODICF SACGVLLNLT VDKDKRVILK EGGGIKKLVD
601 CLRD LGPTDW QLA CLVCKTL WNFSEITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFPKVA QQLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p7, frame 2

TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,
complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete
cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 45/163 (27%), Positives = 77/163 (47%)

Query: 442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSSYYQVK 501
L +++ NI+ H G P VG L + S D+ EE VI T+ NL+ +

Sbjct: 483 LMKMIRNISQHDG--PTKNLFIDYVGDIAAQI---SSDEEEFVIECLGTLANLTIPDL 537

Query: 502 -NSIIQDKKLYIAELLKLLVSNMMDG-ILEAVRVFGNLSQDHDVCDFIQNNVHRFMMA 559
++++ KL + L KL D +LE V + G +S D + ++ + ++

Sbjct: 538 WELVLKEYKL-VPLKDKLPGAAEDDLVEVIMIGTVSMDDSCAALLAKSGIIPALIE 596

Query: 560 LLDAQHQDICFSACGVLL---NLTVDKDKR-VILKEGGGIKKLVDCLRD 604
LL+AQ +D F C ++ + + R VI+KE L+D + D

Sbjct: 597 LLNAQQEDDEF-VCQIIYVFYQMFHQATRDVIKETQAPAYLIDLMDH 644

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 42/178 (23%), Positives = 82/178 (46%)

Query: 169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDISLESLLLEVLRSIDLQTNME 227
K K L V ++ LL V L+ ++ + + + +N +I+ L++ L + N E

Sbjct: 263 KTFKKYQGLVVKQEQLLRVALYLLNLAEADTRTELKMRNKNIVHMLVKALDRD---NFE 318

Query: 228 AFLYCMGSIKFIISGNLGFLENMISKGAVEILINLIKQINENIKKCGTFLPNSGHLVQVT 287
+ + +K +S + N+M+ VE L+ +I +E++ L + +

Sbjct: 319 LLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMPICEHEDL-----LNITLR 366

Query: 288 ATLRLNVDSLSVRKFLNISALPQLCTAM--EQYKGDQDVCT--NIARI--FSKLTYSYRD 341
L D+ L R+K + + LP+L + E YK +C +I+ F + +Y D

Sbjct: 367 LLLNLNLSFDLGL-RNKMVQVGLLPKLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424

Query: 342 CCTAL 346
C L
Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAELLKLLVSNMMDGILEAVRVFGNLSQDHDVCFIVQNNVHRFMMALLDAQHQDICFS 571
I +L+K L +N + ++ V LS + + +V+ ++ ++ ++ +H+D+
Sbjct: 304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDVMEDIVEKLVKMIPCEHEDLLNI 363

Query: 572 ACGVLLNLTVDKDKRVILKEGGGIKKLVDCRLDGLGPTDW-QLACLVCCKTLWNFSENITNA 630
+LLNL+ D R + + G + KL L G ++ Q+A +C L++ S +
Sbjct: 364 TLRLLLNLSFDTGLRNKMVQVGLLPKLTALL--GNENYKQIA--MC-VLYHISMD-DRF 416

Query: 631 SSCFGNEDT-NTLLLLSSFLDEELALD 657
S F D L+ +L DE + L+
Sbjct: 417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 LIFKISRNEKN-DSLIQNDSTLESLLLEVRSE-----DLQTNMEAFLYCMGSIKIFISG 241
LI +++RN N + L+ N++ L +L VLR + +L TN+ +C S G
Sbjct: 155 LILQLARNPDNLEELLNETALGALARVLRREDWKQSVELATNIIYIFFCFSSFSHFHG 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 LVQVTATLRNL----VDSSLVRSKFLNISALPQLCTAMEQYKGDKDVCTNIARIFSKLTS 338
+++ TL NL +D LV ++ +P L ++ +D+ + I S
Sbjct: 521 VIECLGTLANLTIPDLWELVKEY---KLVFPLKDKLKPAAEDDLVLEV-V-IMIGTVS 576

Query: 339 YRDCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGS 398
D C AL + S + L+N Q+ + V +++++ + + R+ KE +
Sbjct: 577 MDDSCAALLAKSGIIPALIELLNAQEDDEFVCQIIYVFYQMFV-HQATRODVIKETQAP 635

Query: 399 QTLTSL 404
L+ L
Sbjct: 636 ATLLIDL 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKKLYIAELLKLLVSNMMDGILEAVRVFGN 537
CE E ++N T + NLS+ ++N ++Q + + L LL + N I A+ V +
Sbjct: 355 CEHEDLLNITLRLLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query: 538 LSQDHDVCD-FIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGGIK 596
+S D F + + + M L + + I +NL +K ++ EG G+K
Sbjct: 410 ISMDDRFKSMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVLICEGNGLK 469

Query: 597 KLVDCLRLDGLPTDWQLACLVCCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEELAL 656
L+ R L D L+ K + N S++ + F + L +SS +EE +
Sbjct: 470 MLMK--RALKLKD---PLLMKMRNISQHDGPTKNLF-IDYVGDLAAQISSDDEEEFVI 522

Query: 657 D 657
+
Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 LNISALPQLCTAM-EQYKGDKDVCTNIARIFSKLTSYRDCTALASYSRCYALFLNLINK 362
LN +AL L + E +K ++ TNI IF +S+ + Y + AL +N+I+
Sbjct: 171 LNETALGALARVLRREDWKQSVELATNIIYIFFCFSSFSHFGLITHY-KIGALCMNIIDH 229

Query: 363 YQKKQDL 369
K+ +L
Sbjct: 230 ELKRHEL 236

Pedant information for DKFZphtes3_14p7, frame 2

Report for DKFZphtes3_14p7.2

[LENGTH] 708
[MW] 79266.35
[pI] 6.57

```

[FUNCAT]      30.25 vacuolar and lysosomal organization      [S. cerevisiae, YEL013w] 3e-04
[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w]
3e-04
[FUNCAT]      09.25 vacuolar and lysosomal biogenesis       [S. cerevisiae, YEL013w] 3e-04
[BLOCKS]      BL00923F Aspartate and glutamate racemases proteins
[BLOCKS]      BL00288B Tissue inhibitors of metalloproteinases proteins
[PROSITE]     MYRISTYL          9
[PROSITE]     AMIDATION         1
[PROSITE]     CK2_PHOSPHO_SITE   12
[PROSITE]     PKC_PHOSPHO_SITE   7
[PROSITE]     ASN_GLYCOSYLATION  11
[KW]          Alpha Beta
[KW]          LOW COMPLEXITY     7.49 %

```

```

SEQ      ESKETVMMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFRGTTSLPSH
SEQ      .....
PRD      cccceeeccccceeeccccccccceeeccccccccccccchhhhhhhcccccccc

```

```

SEQ      LKNGGDQGRHARASSCPSSSDLRLQTKAVPKADLQEEDAEIEVDEVFNTRIVPILRE
SEQ      .....XXXXXXXXXXXX
PRD      CCCCCCchhhhhhhccccccchhhhhhhccccchhhhhhhhhhhhhccccceehhhhhh

```

SEQ LEKEENIETVCAACTQLHHALEEGNMI.GNKFGRSILLKTLCKLVDVGSDSLCLKAKI I
 SEQXXXXXXXXXXXX
 PRD hhhhhcchhhhhhhhhhhhhhhhhhhccccccccccccccchhhhhheeeccccchhhhhhhh

```

SEQ      LALKVSRKNLLNVCKLIFKISRNEKNDSLIQNDSILESLEVLRSDDLQTNMEAFLYCMG
SEG      xxxx.....
PRD      hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhccchhhhhhhhhhhcc

```

SEQ SIKFISGNLGLNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLVQVTATLRNLV
 SEQ
 PRD cccccccccchhhhhhhcchhhhhhhhhhhhhccccccccccccccccceeeehhhhhhhh

SEQ DSSLVRSKFLNISALPQLCTAMEQYKGDKDVCTNIARIFSKLTSYRDCCTALASYSRCYA
 SEQ
 PRD cccchhhhhheeeccchhhhhhhhhhhcccccceehhhhhhhhhhhccchhhhhhhhhhhhh

```

SEQ      LFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSIQTLLSLFQTFHQDLH
SEQ      .....
PRD      hhhhhhhhhhhhhhhhhheeeeeeeccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

```

SEQ SQKPVGQRGEQHRARPPSEAEDVLIKLTRVLANIAIHPGVGPVLAANPGIVGLLLTLE
 SEQ
 PRD cccccccccccccccccccchhhhhhhhhhhhhhhcccceeccccchhhhhhhh

SEQ YKSLDDCEELVINATATINNLSYYQVKSIIQDKKLYIAELLKLLVSNMMDGILEAVRV
 SEGXXXXXXXXXXXXXXXXX.....
 PRD hhhccccchhhhhhhhhheeeccccccccceeehhhhhhhhhhhhhhhhhhhhccccchhhhhhhh

```
SEQ      FGNLSQDHDVCDFI VQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGG
SEG
PRD      ccccccccccccccchhhhhhhhhhhhhccceeeeeecccccccccccc
```

SEQ IKKLVDCRLDGLPTDWQLACLVCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEEL
SEGXXXXXXXXXXXXX
PRD hhhhhhhhcccccccchhhhhhhccccccccccccccccccceehhhhhhhh

```

SEQ      ALDGSFDPDLKNYHKLHWETEFKPAQQLLNRIQRHHTFLEPLPIPSF
SEG      xxx...
PRD      hhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccc

```

Prosite for DKFZphtes3 14p7.2

PS000001	206->210	ASN_GLYCOSYLATION	PD0C00001
PS000001	212->216	ASN_GLYCOSYLATION	PD0C00001
PS000001	311->315	ASN_GLYCOSYLATION	PD0C00001
PS000001	389->399	ASN_GLYCOSYLATION	PD0C00001
PS000001	493->497	ASN_GLYCOSYLATION	PD0C00001
PS000001	500->504	ASN_GLYCOSYLATION	PD0C00001
PS000001	543->547	ASN_GLYCOSYLATION	PD0C00001
PS000001	584->588	ASN_GLYCOSYLATION	PD0C00001
PS000001	628->632	ASN_GLYCOSYLATION	PD0C00001
PS000001	632->636	ASN_GLYCOSYLATION	PD0C00001
PS000001	635->639	ASN_GLYCOSYLATION	PD0C00001
PS000005	173->176	PKC_PHOSPHO_SITE	PD0C00005
PS000005	186->189	PKC_PHOSPHO_SITE	PD0C00005
PS000005	241->244	PKC_PHOSPHO_SITE	PD0C00005

WO 01/12659

PCT/IB00/01496

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_14p7.2)

DKFZphtes3_15a13

group: testes derived

DKFZphtes3_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST

hits

S.cerevisiae Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAAGCGCA  TCGCGTCGG  GCACAGCGG  TGCAGCCTCG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCGAG  TTGCAGAGGA  CTCCCATGAG
151  TGCACCTGGTA  TTCCCAATA  AGATATCAAC  TGAACACCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTTGAGG
251  GGAATATTCC  CAGAATGCGC  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGTCAAAATA  CTGAGAGAAG  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAACCCAG  AAGATCCTCA  GACAATTCCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTTCATAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTTG  TCTACTGACA  CCAAGAAAGC  AAGCATTTCT
551  CTCATTCGCA  AGATTTATAT  CCTAATGCAA  AATCTGGGGC  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGTTATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAAATA
801  TTGACTCAAC  TATACTATCA  CCAAAACAAA  TAAAAACACC  ATTTCAAAAA
851  ATCTGAGGG  ACAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTTGGAC  ATTGAACTA  AAATGGAAGA  ACAGGAAAAA  AACCTGCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAATTATG
1001  AGGTCTAAAG  AAAATCCAGA  TCTTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAATAGT
1201  CCTCCATCAC  TTTGATTCTT  CTAGTCAAGA  GTCAAGTCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTTGCAGAGT  TCTTCTCACC  ATTTAAACTG  AAGGACCCCTA  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACACTAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATAA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTC
1501  GGTAATAAGT  AAAATTTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCCTTTAAA  TATATACTAA  CTGTAAAGGC  CCTCTAATGC  CATTTTTCTA
1601  AACAGTAATG  TTTACTTTGG  TATTAAATTT  TGGTATGGAT  TCACCTTTTA
1651  CTTATGTTAA  AATTATACCA  TTTAACTGGC  TCTTTGTGCA  TTGTGCTGTT
1701  ATTAACACAA  TGTCTTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAGG
1801  GGCGGCGCT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387
Category: similarity to known protein

```
1 MATAQLQRTP MSALVFPNKI STEHQSLVLV KRLLAVSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDK NCPGSTQLVK WMLGCDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNNGLP MDFISKQSN ESSMLSTDTK KASILLIRKI
151 YILMQNLGPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERMENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQEKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLISHSQV EQLVKNKSEL DMESKTRSG KVFQNMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSSS QESVPKRRKF SEPKEHI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15a13, frame 2

TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.
Length = 562

HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22
Identities = 84/290 (28%), Positives = 145/290 (50%)

```
Query: 22 TEHQSLVLVKRLLAVSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSTQLVKW 81
TE SL+L + LL +++ I+Y+RG+FPE + + + L +KI + S +L+ W
Sbjct: 11 TEQDSLLLRNLLRIAIFNISYIRGLFPEKYFNOKSVPALDMKIKKLMFMDAESRRLIDW 70

Query: 82 M-LGCDYDALQKKYVYT-----NPEDPQTISECYQFKFYTNNGP--LMDFISK--NQSN 130
M G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct: 71 MEKGVYDALQRKYLKTLMFISICETVDGPMIEE-YSFYSYSDSDSQDVMNINRTGNKKN 129

Query: 131 ESSMLST-----DTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPDYQPP 184
ST + ++ ++R + LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct: 130 GGIFNSTADITPNQMRSSACKMVRTLVQLMRTLDMKMPDERTIVMKLLYYDDVTPPDYEP 189

Query: 185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVT-----ERERMENIDSTILS 235
F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct: 190 FFRGCTEAEQYVWTKNPLRMEIGNVNSKHLVLTLLKVKSVLPDCEDEDDMDQD-DGKSIG 248

Query: 236 PKQIKTPFKILRDKDVEDEQEHY-----TSDDLDIETKMEEQEKNPASSE 281
P + Q D ++ QE+ DD D E ++ +PA +E
Sbjct: 249 PDSVHDD-QPSDSDSEISQTQENQFIVAPVEKQDDDDGEVDEDDNTQDPAENE 300
```

Pedant information for DKF2phtes3_15a13, frame 2

Report for DKF2phtes3_15a13.2

```
[LENGTH] 387
[MW] 44417.64
[pI] 5.57
[HOMOL] TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 03.13 meiosis [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, Y1L072w] 7e-11
[PIRKW] nucleus 2e-09
[PIRKW] zinc finger 2e-09
```

```

[PIRKW]      DNA binding 2e-09
[PROSITE]    MYRISTYL 1
[PROSITE]    CAMP_PHOSPHO_SITE 3
[PROSITE]    CK2_PHOSPHO_SITE 12
[PROSITE]    PKC_PHOSPHO_SITE 7
[PROSITE]    ASN_GLYCOSYLATION 3
[KW]         Alpha_Beta

```

```

SEQ  MATAQLQRTFMSALVFFNKISTEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDD
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ  LCVKILREDKNCPGSTQLVWMLGCVDALQKKYVYTNPEDPQTISECYQFKFYTNNGPL
PRD  hhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccce

SEQ  MDFISKQSNESMLSTDTKKASILLIRKIYILMQNLGFLPNDVCLTMKLFYDEVTPPD
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  YQPPGFKDGDCEGVI FEGEPMYLVNVEVSTPFHIFKVKVTTERREREMENIDSTILSPKQIK
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ  TPFQKILRDKDVEDEQEYHTSDDDLOIETKMEEQEKNPASSELEEPSLVCEEDEIMASKES
PRD  hhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ  PDLISHSQVEQLVNKTSSELDMSSEKTRSGKVFNKMANGNQPVKSSKENRKRKSQESGR
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ  IVLHHFDSSSQESVPKRRKFSEPKHI
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15a13.2)

DKFZphtes3_15c24

group: metabolism

DKFZphtes3_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatGG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
1 CGAAGGCGGC GCGAAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGAAC TGCCCAGGAG AGGAGTCTGC AGGTCCCAG GAGCGGCGAC
151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATGTGAA
251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTGAAG TACACAATA
501 TAATATAACC ACAGTGGAAA ACTTTCACAA TTTCATGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTGTA ATGAACCTGG
651 ACAAAACATGG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGCCTG TGCTCCACCA
751 CTTGTAGTTG CTGCAAAAT TGAATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTGTGCAGCC AGTCTTCCTA CCACATGCGG TGTGGTTGCT GGGATCTTAG
851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTGTGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCAGCT CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAGAG GAACTGAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATAC ACAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGAAGA TTCTGGTGAA AGCTTGGAAG ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAACT TGATAAAACT
1351 TAGGGCAACA TTAATTAAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAACTCTG TGACTTGCTT GTTTCTCCCC GCTCCAACGA AATCATTAACT
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTT TTTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTTCTATAAG AAAATGCCCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACTCG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAATAAAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAATAA
1951 AAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404
Category: similarity to unknown protein.
Classification: Metabolism
Prosites motifs: D_2_HYDROXYACID_DH_1 (76-105)

```
1 MAESVERLQQ RVQELERELA QERSLQVPRS GDGGGGRVRI EKMSSEVVDS
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LLFDYDKVEL ANMNRLLFFQ HQAGLSKVQA AEHTLRNINP DVLFEVHNYN
151 ITTVENFQHF MDRISNGGLE EGKPVDLVLS CVDNFEARMT INTACNELGQ
201 TWMESGVSEN AVSGHIQLII PGESACFACA PPLVVAANID ETKLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPTMSMKPN
301 PQCDNRNCRK QEEYKKKVA ALPKQEVQEE EEEIHEDNE WGIELVSEVS
351 EELKNFSGP VPDLPegITV AYTIKKQED SVTELTVEDS GESLEDLMAR
401 MKNM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15c24, frame 1

TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid
T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1
YAC YUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus
fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796_4 gene: "moeB"; product: "MoeB"; Staphylococcus
carnosus molybdenum cofactor biosynthetic gene cluster, complete
sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.
Length = 419

HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122
Identities = 241/367 (65%), Positives = 293/367 (79%)

```
Query: 37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96
      R +IEK+S+EVVDSNPYSRLMAL+RMGIV++YE+IR VA+VGVGGVGSV AEMLTRCG
Sbjct: 48 RQKIEKLSAEVVDSNPYSRLMALQRMGIVNEYERIREKTVAIVGVGGVGSVVAEMLTRCG 107

Query: 97 IGKLLLFYDKVELANMNRLLFFQHQAGLSKVQAAEHTLRNINPDVLFVHNYNITTVEN 156
      IGKL+LFYDKVE+ANMNRLLF+Q+QAGLSKV+AA TL ++NPDV EVHN+NITT++N
Sbjct: 108 IGKLLLFYDKVEIANMNRLLFYQPNQAGLSKVEAARDTLIHVNPDVQIEVHNFNITMDN 167

Query: 157 FQHFMDRISNGGLEEGKPVLDVLSVDNFEARMTINTACNELGQTMESGVSENAVSGHI 216
      F F++RI G L +GK +DLVLSVDNFEARM +N ACNE Q WMESGVSENAVSGHI
Sbjct: 168 FDTFVNRIKGSLLTDGK-IDLVLSVDNFEARMVNMACEENQIMESGVSENAVSGHI 226

Query: 217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLFKLLNF 276
      Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTTM VVAG LV N LK+LLNF
Sbjct: 227 QYIEPGKTACFACVPLVVASGIDERTLKRKGVCAASLPTTMVAVAGFLVMNTLYLLNF 286

Query: 277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDNRNCRKQEEYKKKVAALPKQ-EV-IQEEEEEI 334
      G VS Y+GYNA+ DFFP S+KPNP CDD +C ++Q+EY++KVA P EV +EEE +
Sbjct: 287 GEVSQVVGYNALSDFPRDSIKPNPYCDDSHCLQRQKEYEEKVANQPVDLEVEVPPEETV 346

Query: 335 IHEDNEWGIELVSEVSEELKNFSGPVPDLPegITVAYTIKKQEDSVTELTVEDSGESL 394
      +HEDNEWGIELV+E SE + S + G+ AY P K+ D+ TEL+ + +
Sbjct: 347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAIE-PIKR-DAQTELSPAQA--AT 399

Query: 395 EDLMAMKKN 403
```


Pedant information for DKF2phtes3_15c24, frame 1

Report for DKFZphtes3 15c24.1

```

SEQ      MAESVERLQORVQELERELAQERSLQVPRSGDGGGGRVRIEKMSSEVVDNSPNYSRLMALK
SEG
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhc
MEM      .....

SEQ      RMGIVSDYEKIRTFAVAIVGVGGVSVTAEMLTRCGIGKLLFDYDKVELANMNLFFO
SEG      .....xxxxxxx.....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhc
MEM      MCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      HQAGLSKVQAAEHLTARNINPDVLFVHHYNIITTVENFQHFMDRISNGLEEKGPDVLVLS
SEG
PRD      cccccchhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhccccccccccceeee
MEM      .....

SEQ      CVDNFEARMTINTACNELGOTWMESGVSENAVSGHIQLIIPGESACFACAPPLVVAANID
SEG
PRD      cccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      EKTLRREGVCAASLPTTMGVAGILVQNVLRFLNLFGTVSFYLYGNAMQDFPPTSMKPN
SEG
PRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccc
MEM      .....

SEQ      PQCDRRNCRKQOEYKKKVAALPKQEVIEEEEEIHEDNWGIELVSEVSEELKNFSGP
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXX.....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhc
MEM      .....

SEQ      VPDLPegITVAYTIPKQEDSVTLETVEDSGESLEDLMAKMKNM
SEG
PRD      ccccccccccccccehhhhhhhhhhhhhhccccccccchhhhhhhhhhhccc
MEM

```

Prosite for DKFZphtes3 15c24.1

(No Pfam data available for DKFZphtes3_15c24.1)

DKF2phtes3_15c6

group: transmembrane protein

DKF2phtes3_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```
1 GAGACACTGA GCCCCGAGAC AGTGAGTGGT GGCCTCACTG CTCTGCCCGG
51 CACCCCTGTCA CCTCCACTTT GCCTTGTGG AAGTGACCCA GCCCCCTCCC
101 CTTCCTTCT CCCACCTGTT CCCAGGACT CACCCAGCC CCTGCCCTCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGGCTTGGAG CCAGAGACAC
251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGGCCGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCCTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCCTGGCTA GGAGTGTTCG
401 TGTTTCCTCC TTTGCCCTACC ACTCTGGGGT GGGCAGTGT GTGGGGAAGC
451 TGGCTGTGGG ATGGTAGCTA TTCCACCTC TGCCTGCCTG CCTGCCCTGT
501 GTCTTGGGCA TGGTGCAGTA CCTGTGCCCTA GGATTGGTTT TAAATTTGTA
551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGAAGTCTCT
601 TCCACAGCC TGCGCTTGCC TCCCTGCCCTC ATCTCTATTC TCATTCCACT
651 ATGCCCCAAG CCTGGTGGT CTGGCCCTTT CTTTTCTCTC CTATCCCTAG
701 GGACCTGTGC TGCTCTGCC TCATGTCCCA CTGGTGTGT TAGTTGAGGC
751 ACTTTATAAT TTTCTCTTG TCTTGTGTT TTTCTGCTT TATTTCCCTG
801 CTGTGTCTCT TCCTTAGCAG CTCACCCCA TCCTTTGCCA GCTCCTCCTA
851 TCCCGTGGG ACTGGCCAAG CTTTAGGGAG GCTCCTGGT TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGT GGTGAGGCA GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACTTC ACTGCATCCT TGCCCCATTC AGCCCGGCT
1001 TTATGATGAC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGGTGA TCCAAGTGA GTGGGACCCC TACTAGGGT CAGGAAGTGG
1151 AACTAATCAT CTGTGCAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGGCGGCC GCTCTAGAGG
1251 ATCCAAGCTT ACCTAAAAAA AAAAAAATAA AAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2
-----ORF from 461 bp to 814 bp; peptide length: 118
Category: putative protein

```
1 MVAIPPSACL PACCPGHGAV FVPRIGFKFV NNFFFLVDV NRAREVLPTA
51 CACLASSSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCFLL YFPAVSCP
```

No BLASTP hits available

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76. P = 0.33

HSPs:

Query: 15 PGHGAVPVPRIGFKFVNFFPFLGVDVNRAREVLPTACACLPASSLFSFHYAPSPGGLALS 74
PG GA P+ R+ F+ P+ +E+ A C P SSL+ A G L
Sbjct: 52 PGRGA-FLARVTRFH----PFF---KKQELFVAAEVCTPVSSLYCGKKATLVVGN 103

Query: 75 FSSYPQGPVLLCP---HV-PLGCLVEALYNFSLVL 105
S P+G V+ C HV G L A +++V+
Sbjct: 104 LRSIPEGAVV-CNVEHHVGDGVGLARASGDYAI 137

Report for DKFZphtes3 15c6.2

```
SEQ      MVAIPPSACIAPCCPGHGAVPVPRIGFKFVNNFPGFLVDVNRAREVLPTACALPASSLF
PRD      cccccccccccccccccccccccccceeeccccceehhhhhhhcccccecccccccc
MEM

SEQ      SFHYAPSPGGLALSSSYPOGVLVLCPHVPLGCLVEALNFSVLVCSFLLYFPAVSCP
PRD      eccccccccceeeccccccccccccccccchhhhhhhhhchhhhhhhcccccccc
MEM      .MMMMMMMMMMMMMMMMMMMM
```

PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE ZIPPER	PDOC00029

(No Pfam data available for DKF2phtes3_15c6.2)

DKFZphtes3_15g14

group: testes derived

DKFZphtes3_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```
1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATTG TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTTAGAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACTT CCAATGTGA AGAAAAAGCT
401 GATGTTTTAA GCTCCTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCCTGT GATGTAAGAG AGAAGTGGCT TTCTAAAAA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAGAGG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCATTTT TAGTAACGTG
601 AGGAAAAAAC AGTGAAATTG TTGTAAACC AAATCTTGAA TATAAAGAAC
651 TTTGTCATTT GGTATCTGAA GAGGAAGCAT TTGACTTTTT TAAATATTTG
701 GATGCAAGA AAGAAAATTC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCACAGA AAGCTGTGCC ACCATTTTGT CAACAAAAAG TTTGAAACCC
801 TTTGGAAGAC CAAATCTTTT TCTAAATGA ATTGCATGTC TGGTAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCCGGAA AAGCAGACA AACGTGGGAA
901 AAGGCCCTCT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTACGAAA GGAACACCTG GAAATGTTTG AAGCGATTGG TTTTITAGCT
1001 ATCAAACTTG GTGTTATTCC TTCGGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCATATC AAGCAATGGT TGTAGAAAA GTGACTCCAG
1101 AGAGGTTGAA AATATTGAA AAGAAATTG AAAAGAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATCCCTGT AGACTTGGTC AGCTCAAGG
1201 AATCACTTTT GATATTGTCA TTAGAAATTT AAAAAACAA ATAAATGATT
1251 CTGCAAACTT GAGGGAGAGA ATTATGGAA CAATAGAAAA TGTAAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAAATT GACTAGCTTT GCTGAAGAA GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAAG ACTTGGATGA TCCTGTAAT
1451 AGAGCAAGA AGTATTTTCT TCAAACTGAG GATGCTAAG GCACACTTTC
1501 ATTGATGCCT GAATTCAAAG TCCGTGAGAG AGCATTTGTT GAGGCATTGC
1551 ACCGCTTTGG CATGACCAGG GAAGGTTGTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTCCA TGCGCATATT CTATGTTTAC GCATATACCA GCAAAATTTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCAGTACC
1851 CGAAGACAA AGTAGGGCAG TGGTACCATT ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTGGTATAA CCATATATAT GTCAACCTTT CTTGTTTTTG AAATATTGTA
2201 TCAGAACAA ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTATAT CTGGCATACT GCTGACCTTC
2351 TTGACCATTT GTAATTTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCAT TCTTAGATTG TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTGAAAAA TTTTAATTTT TGTGTTGTTA
2601 ACTATCCAC ATTATAAATC ATCCTTCACC ATTTATATC AGTTAAATAT
2651 GGGTGTGTTG GGGAGGAATG ACTGCCATGT AGACATGTAT TGATTTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAATTTAA
```

```

2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAACC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTG CGTAACATAA
2951 GTAAGTAAAA TGATGTGTAT CAACTTGATG GTAAAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTTACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAT AGCACAGATA TTGGAGACAA
3151 ACTAACCAG TTTGAACCCCT GGCACCTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAGATTAT TTAACCTCAT GGGCTTCAGT TTCAACATCT GTAAAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAT TTTCTAAGTT
3301 AATATATGTA AAGCACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAAA TAAATGAAAA
3401 TCCCTTCTGT TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701
 Category: similarity to unknown protein

```

1 MEEDTDYRIR FSSLCFFNDH VGPHGT:KSS PSDFIVIEID EQQQLVNKTI
51 DEPIFKISEI QLEPNNFPPK PKLDLQNLSL EDGRNQEVHT LIKYTDGDDQ
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNMF ACQVREKWL
151 KTELIGLPPE FSGIRILOKN QRASLHSAIR QKFPFLVTVG KNSEIVVKPN
201 LEYKELCHLV SEEEAFDFFK YLDAKKENSK FTFKPDNTKD HRKAVHHFVN
251 KKFGNLVETK SFSKMNCAG NPNVVVTVRF REKAHKGKRR PLSECQEGKV
301 IYTAFTLRKE NLEMFEAIGF LAIKLGVIPS DFSYAGLKDK KAITYQAMVV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKDN HFDIVIRNLK
401 KQINDSANLR ERIMEAIENV KKGKGVNYYG PORFGKGRKV HTDQIGLALL
451 KNEMMKAIKL FLTPEDLDDP VNRKKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHFRFG TEEGCIQAWF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDDIDEN FPNSKIHLVT EEEGSANMYA IHQVLPVPLG
601 YNIQYPKNKV GQWYHDILSR DGLQTCRFKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTKGSHI DETALSLTIS FDLASCYAT VCLKEIMKHD
701 V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15g14, frame 2

TREMBL:SPBC1A45P_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid c1A4 left hand region 1-26184 bp
 Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)
 Length = 676

HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
 Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSEIVKPNLEYKELCHLVSEEEAFDFFK-YLDAKKENSKFTFKPDNTKDHRAVHHFV 249
 + E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPRLR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNLVETKSF SKMNC SAGNPVVTVRPREKAHK-RGKRPLSECQEG-KVIYTAFTL 307
 + F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNSRTNKQEKINQTRDANGVENWGYGPKDFIHFTL 224

Query: 308 RKENLEMFEAIGFLAIKLGVIPSD-FSYAGLKDKKAITQAMVVRKVTPERLKNIEKEIE 366
 KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVISIKIGLDRNLNLRNL- 282

Query: 367 KKRMMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKGFV 426
 K M + N D SL LG LKGN F +VIR++ N +L E + +++ + GF+

Sbjct: 283 -KGMIIIGNYNFSDASLNLGDLKGNFVIVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGQRFQGRKRVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNR-AKKYFLQTEDAK 485
 NY+G QRFQ + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFQGT-F-SISTHTIGRELLSNWKKAAELILSDQDNVLPKSKARKIWAETKDAA 399

Query: 486 GTLSLMPEFKVRERALLEALHRFGMTEEGCIQ--AWFS----LPHSMRIFYVHAYTSKIW 539
 L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPROCLAENALLYSLSNQRKEEDGTYSENAYYTAIMKIPRNLRTMYVHAYQSYVW 459

Query: 540 NEAVSYRLETYGARVVQGDVLC-----LDEDIDDENFPNS-----KIHLVTEEGS 585
 N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NSIASRIELHGLKLVVGDVVIDTSEKSPGISDDEDFEDVREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVVLPLVGYNIQYFKNK-VGQWYHDILSRDGLQTCRFKVPVTLKLNPGCYRQI 644
 + Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLPSPGFDVLYPSNEELKQLYVDILKADNMDPFNMRRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671
 ++ P +L Y+++ D + + +D

Sbjct: 580 IQPKSLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01
 Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFVIEIDEQQQLVNKTIDEPIFKISEIQLEPNFPPKPKLDLQNLSE 81
 GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIKYPDGDQNHQSGS--EKEDTI-VDGTSKCEKADVLSSFLDEKTHELLN 138
 R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKIH 161

Query: 139 NFACDVREKWLKTELGIPPE-FSIGRILDKNQASLHSAIRQ 181
 +RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
 Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698
 ++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKFLQLGTSAYATMALRELK 660

Pedant information for DKF2phtes3_15g14, frame 2

Report for DKF2phtes3_15g14.2

[LENGTH] 701
 [MW] 80700.96
 [PI] 7.31
 [HOMOL] PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae) 2e-51
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53
 [BLOCKS] BL01268C
 [BLOCKS] BL01268B
 [BLOCKS] BL01268A
 [SUPFAM] hypothetical protein HI0701 3e-06
 [PROSITE] MYRISTYL 7
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 16
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 13
 [PROSITE] ASN_GLYCOSYLATION 5
 [KW] Alpha_Beta

```

SEQ MEEDTDYRIRFSSLCFFNDHVGFGHTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEI
PRD ccccccccccccccceccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QLEPNFPKKPKLDLQNLSDLEGRNQEVHTLIKYTDGQDNHQSSEKEDTIVDGTSKCEE
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ KADVLSSFLDEKTHELLNNFACDVREKWSKTELIGLPPEFSIGRILOKNQRASLHSAIR
PRD hhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhheccccccccccccccccccccchhhhhhhhhhh

SEQ QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEAFFDKYLDACKENSKFTFKPDNTKND
PRD hccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhcccccceccccccccccch

SEQ HRKAVHHFVNKKFGNLVETKSFSKMNCSSAGNPVVTVRFREKAHKGKRPSECEQGV
PRD hhhhhhhhhhhhhhhhhhhheccccccccccccccccccccchhhhhhhhhcccccccccccccc

SEQ IYTAFTLRKENLEMFAGFLAIGLVI PSDFSAGLKDKKAITTYQAMVVRKVTPERLKN
PRD hccccccccccccchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhheccccccccchhhhh

SEQ IEKEIEKKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENV
PRD hhhhhhhhhhhheccccccccccccccccccccccccccccccccchhhhhhhccccchhhhhhhhhhh

SEQ KKGKGVNYYGQPRFGKGRKVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNRKAYFLQ
PRD hccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhh

SEQ TEDAKGTLMLPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRI FYVHAYTSKIWN
PRD hccccchhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhh

SEQ EAVSYRLETYGARVVQDGLVCLDEDIDDENFPNSKIHLVTEEGSANMYAIHQVVLVPLG
PRD hhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ YNIQYPKNKVGQWYHDILSRDGLQTCRFKVPVLKLNIPGCRYQLKHPCNLSYQLMEDHD
PRD cccccccccchhhhhhhhhhhccccccccccccccccccccccccchhhhhhhccccchhhhhhhhhhhc

SEQ IDVKTGSHIDETALLISFDLDASCIATVCLKEIMKHDV
PRD cccccccccchhhhhhhhhheccccccccchhhhhhhhhhhcccc

```

Prosites for DKFZphtes3_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphes3_15g14.2)

DKFZphtes3_15h1

group: testes derived

DKFZphtes3_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACAGATA GAGTTCTCC AGCTTTTCTT TGATTGCTC TGCTTTAGCG
51 TCTCTAAATC CGGTACCAT GTCGGACCCC GAAGGCGAGA CTTGCGAAG
101 CACCTTTCCC TCTTATATGG CCGAAGCGGA GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCTCGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTCTGT TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCGAG GAAAGCATCA
451 ACAACTCAGT GGAAGTCCT TCTTCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAAT ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATG AACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGGA
601 AGGCCCTCGT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGAGCTC
651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGCGC GCCTGACTGT GGAGGACCTC ATCATGACCG
751 GCATCAACTA CCTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTCGG
851 GGACCACAAA CGCCGTCCTT CACAGACAGC CCATTACATC CTCGAAGGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGCTTT
951 CAGAAAGCTG AGAAAGTGCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAC AAGGATGAAC TGGTTGGAAA CTTGTATAGC TGCAATAGGA
1051 ATGCCAGAT TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAGATCCC CTCTGGCAAA AACCACCTTG
1251 GAGAAGACCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCTCGG CAGGCCGAGA ATTATGGCGA GAAGTCCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGACCAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGGTATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAAA GAAAAAAGCG AGGGAGAAGC
1601 TTCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCGAGAAG GTGCTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAAGGCC CTGTGAGGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CTTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAAGAAAC TTTCAGAAAT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAATTT GGAGAAATAG GAGAAACGAA
2051 AAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTTC AAGCGATTGG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACTCTT
2251 GCAAAAAATA AAAAAAATA AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672
Category: similarity to known protein

1	MSDFEGETLR	STFPSYMAEG	ERLYLCGFEF	KAGJQFSNAL	LYQDQDFEAL
51	VARSKQKMLN	GLDERSLKDA	EASLQSDPAF	CKKQSLQKAT	LYTMDGDFCA
151	LVTHFGKXKJ	QKQKQKQKQK	QKQKQKQKQK	VGPSPSKICE	NKGLDVSFLSK
251	QAKN	POPMKLLHFF	QKQKQKQKQK	LKSKETVQRL	LCGLYVDQKE
351	KKLLKDDLEL	IKGTMKRGSL	VEDLMTGTIN	LYDHTSNFVR	QKKPIYARER
451	KKLLKDDLEL	QKQKQKQKQK	QKQKQKQKQK	LYDMLTSGSA	QKQKQKQKQK
551	KKLVKNEEVL	EVPNKDELVG	ALYSCIGNAQ	LYDQGEAAAL	QSHRKKDLTA
651	KEYOLDPAKS	RALDNIQGRV	NRVGCFOQAI	OTWEKCIPLA	TKTLTKTWLF
751	HEIGRCYLKE	DQAWAQQKQK	EKSQOQCAEE	QDEIOWNLAS	VLVROAQVRL
851	RDFAVSANNF	EKALERAQKV	HNNEAQQAII	SALDDANKGI	IRLRLKNAQV
951	ENLNKSEKSE	ASLYEDRIT	REKDMRRVRD	EPEKVQKWMD	HSEDEKTECI
1051	DDEAFGEALQ	SPASQKQSGE	AGKARSDLGA	VAGLSGKSEL	TRSGETGRKL
1151	LEAGRSESES	IYRRPSGELE	QRLSGEFSRQ	EPEELKKLSE	VGRREPEELG
1251	KTOFGEIGET	KKTNGNEKE	ME		

BLASTP hits

Entry AF039202.1 from database TREMBL:
product: "Hsp70/Hsp90 organizing protein"; *Cricetulus griseus*
Hsp70/Hsp90 organizing protein mRNA, complete cds.
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782_1 from database TREMBL:
product: "myosin heavy chain"; Argopecten irradians myosin heavy chain
mRNA, complete cds.
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:
stress-induced protein stil - soybean
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3.15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15h1, frame 3

Report for DKFZphtes3_15h1.3

```

[LENGTH]          672
[MW]               76655.61
[pI]              5.49
[HOMOL]           PIR:S56658 stress-induced protein stil - soybean 6e-10
[SUPFAM]          tetratricopeptide repeat homology 1e-07
[PROSITE]         MYRISTYL      7
[PROSITE]         AMIDATION     3
[PROSITE]         CAMP_PHOSPHO_SITE      4
[PROSITE]         CK2_PHOSPHO_SITE      15
[PROSITE]         TYR_PHOSPHO_SITE       1
[PROSITE]         PKC_PHOSPHO_SITE      11
[PROSITE]         ASN_GLYCOSYLATION      2
[KW]              All_Alpha
[KW]              LOW_COMPLEXITY      4.76 %

```

```

SEQ      MSDPEGETLASTFSPSYMAEGERLYLCGEFSKAAQSFSNALYLDQDGKNCILVARSKCF LKM
SEG
PRD      cccccccceccccccccccccccccchhhhhhhhhhhhhhhhhccccceehhhhhhhhhhh

SEQ      GDRLSKDAEASLQSDPAFCGILQKATLTMTGDEFAFVHYRHYKLRPDREFRVGI
SEG
PRD      hcchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccchhhhhh

```

```

SEQ      QKAQEAINNVSVPSSIKLENGKDLFSLSKQAEINIAKQKQPQPMKHLHPTKGEPKWKA
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhcchhhhhhhcccchhhhhcccceccchhhh

SEQ      LKSEKTVRQLLGELYVDKEYLEKLLDLEDLIGKTMGGTLVEDLIMTGINYLDTHSNFR
SEG
PRD      . . . . . xxxxxxxxxxxxxxxxxxxx . . . . .
          hhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhcccceccccc

SEQ      QOKPIYARERDRKLQEKWLDRDHRKRPQTAYHYLLSKLEDIDLMLTSGSAEGSLQAKEV
SEG
PRD      . cchhhhhhhhhhhhhhhhhhhhhcccceccchhhhhhhhhhhhhheeececcchhhhhhhh

SEQ      LKKVLEWKEEVPNKDELVGNYLSCIGNAQIELQGMAEALQSHRKDLIAKEYDLPKAS
SEG
PRD      . hhhhhhhcccceccceeececcceccchhhhhhhhhhhhhhhhhhhhhhhhhhhcccchh

SEQ      RALDNI GRVFAVKGQQAIDTWEKII PLAKTTLEKTWLFHEIGRCYLELDQAWQAQNYG
SEG
PRD      . hhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhh

SEQ      EKSQQCAEEEGDIEWQLNASVLVAQAQVKLRDFESAVNNFEKALERAKLVHNNEAQAQAI
SEG
PRD      . hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhh

SEQ      SAlDDANKGI IRELKRTNYVENLKEKSGEASLYEDRI I TREKDMRVRVDEPEKVVQKD
SEG
PRD      . hhhcccchhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhcccceeecc

SEQ      HSEDEKTEDDEAFGEALQSPASGQKQSV EAGKARSDLGAVAGLSGELGTRSGETGRKL
SEG
PRD      . xxxxxxxxxx . . . . .
          cccccccccchhhhhhhhhcccceccchhhhhhhccccecccecccecccccchhh

SEQ      LEAGRRRESREIYRPSGLEQRLSGEFSQEPFELKKLSVEGVRREPELGGKTFGEIGET
SEG
PRD      . hhhcccceeececccccchhhhhcccceccchhhhhhhhhhhhhhhhhcccceccccc

SEQ      KKTGNEMKEEYE
SEG      . . . . .
PRD      . cccccccccccc

```

Prosites for DKFZphtes3_15h1.3

PS000001	128->132	ASN_GLYCOSYLATION	PDCC000001
PS000001	438->442	ASN_GLYCOSYLATION	PDCC000001
PS000004	265->269	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	605->609	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	613->617	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	636->640	CAMP_PHOSPHO_SITE	PDCC000004
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	66->69	PKC_PHOSPHO_SITE	PDCC000005
PS000005	136->139	PKC_PHOSPHO_SITE	PDCC000005
PS000005	180->183	PKC_PHOSPHO_SITE	PDCC000005
PS000005	183->186	PKC_PHOSPHO_SITE	PDCC000005
PS000005	186->189	PKC_PHOSPHO_SITE	PDCC000005
PS000005	214->217	PKC_PHOSPHO_SITE	PDCC000005
PS000005	342->345	PKC_PHOSPHO_SITE	PDCC000005
PS000005	564->567	PKC_PHOSPHO_SITE	PDCC000005
PS000005	596->599	PKC_PHOSPHO_SITE	PDCC000005
PS000005	660->663	PKC_PHOSPHO_SITE	PDCC000005
PS000006	2->6	CK2_PHOSPHO_SITE	PDCC000006
PS000006	66->70	CK2_PHOSPHO_SITE	PDCC000006
PS000006	93->97	CK2_PHOSPHO_SITE	PDCC000006
PS000006	171->175	CK2_PHOSPHO_SITE	PDCC000006
PS000006	220->224	CK2_PHOSPHO_SITE	PDCC000006
PS000006	277->281	CK2_PHOSPHO_SITE	PDCC000006
PS000006	382->386	CK2_PHOSPHO_SITE	PDCC000006
PS000006	392->396	CK2_PHOSPHO_SITE	PDCC000006
PS000006	481->485	CK2_PHOSPHO_SITE	PDCC000006
PS000006	507->511	CK2_PHOSPHO_SITE	PDCC000006
PS000006	512->516	CK2_PHOSPHO_SITE	PDCC000006
PS000006	542->546	CK2_PHOSPHO_SITE	PDCC000006
PS000006	548->552	CK2_PHOSPHO_SITE	PDCC000006
PS000006	628->632	CK2_PHOSPHO_SITE	PDCC000006
PS000006	663->667	CK2_PHOSPHO_SITE	PDCC000006
PS000007	506->515	MYRISTYLSITE	PDCC000007
PS000007	119->125	MYRISTYL	PDCC000008
PS000008	132->138	MYRISTYL	PDCC000008
PS000008	213->219	MYRISTYL	PDCC000008

WO 01/12659

PCT/IB00/01496

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15h1.3)

DKFZphtes3_15i5

group: cell structure and motility

DKFZphtes3_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)
"radial spokehead" part of flagella in *Chlamydomonas*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```

1  CACCCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAAACGGCC CCCTCTCTCG
51  GTGCTCAGAA ACCGGCCGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101  TCTGCAAGCC TTTCTCTTAG AGATCTCTGC CTCTGGGGA ACCATGGGAG
151  ACCTGCCGCC CTACCTGAGC CGCCCTGCCC AGCAGCTCC GGGCCGGAGG
201  ACTTCTCAGG CTTCCACAGG GCGGCACAGT CGGACCAAG CTCAGGCCCT
251  GGCAGCGGAG CCGAGAGAGA GCGAGCAGAT ACCTCCAGAC GCCACGGAG
301  AGCCCTCTGG TTGGTCACAG AGGGGACAGC TGTCCCAACA GGAGAAGTTG
351  CTGATGGCCCC AGGTCTTCCA GCTTGACGAA GCGCCGCTGG GTGCCATGGA
401  GTACCCATCT GTGAACACGG GCTTCCCTC AGAGTCCAG CCTCAGCCTT
451  ACTCTGATGA AAGCAGGATG CAGCTCCGCG AGCTCACCAC CAGCCTAATG
501  CTGCAAGCGG TCCAGCAGGG CCAAGCAGCG CTGTTCCAGC AACTGGACCC
551  CACCTTCCAG GAGCCCCGAG TCAACCCCTT GGGCCAGTTC AACTCTTACC
601  AGACAGACCA GTTCTCTGAA GGTGCCAGC AGGGGCTTCA CATAAGGGAT
651  GACCCCTGCC TTCACTCTTT GCGCTCTGAG CTGGGCTTCC CACACTACAG
701  TGCCCAAGTG CTTGAGCCCG AGCCTCTGGA GCTGGCCGTG CAGAAGCCCA
751  AGCCCTACCT GCTGCAGACC AGCATCAATT GCGACCTCAG CCTGTACGAG
801  CACTTGGTAA ATCTGCTGAC CAAGATCCTG AACCAAGCGG CTGAGGACCC
851  CTGTCTCTCT CTGAGCTCTC TGAACCGCAC CACCGAGTGG GAGTGGTTCC
901  ACCCCAGACT GGACACGCTG CGGACGACCC CCGAGATGCA CCCACCTAC
951  AAGATGGCGG AGAACAGAA GCGCTCTTTC ACCCGAGTG GAGGCGGCAC
1001  TGAAGGCGAA CAGGAGATGG AGGAGCAGGT GGGGGAGACA CCAAGTGCCA
1051  ACATCATCGA GACTGCCTTC TACTTCGAGC AGGCCGGGCT CGGCTGAGC
1101  TCGGACGAGA GCTTCCGCAT TTTCTGGGCC ATGAACAGC TGGTGGAGCA
1151  GCAGCCCATC CACACCTCTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201  GCAGCTACCT GGTGCCCGAG GTGGAATTCC GGGAGGGGGA GGAGGAGGCA
1251  GCAGAGGAGG AGCTGCAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301  GCAGGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351  TCCCTAAGTC CGTATGGAAG CCGCCGCCCG TGATCCCAAA GGAGGAGAGC
1401  CGCTCAGGCG CCAACAAGTA CCTGTACTTT GTGTGCAAGC AGCCGGGCCT
1451  GCCATGGAGC CGGCTCCCCC ACCTCATTCC AGCCAGATGT GTGACGCCCC
1501  GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACAGGCC AGTCGTGAGC
1551  TACCCAGCCT TCCCGGGCAA CGAGGCCAAC TACCTCGGGC CCCAGATAGC
1601  CCGCATCTCG GCCGCCACGC AGGTACAGCC GCTGGGCTTC TACCAGTTTA
1651  GTGAGGAGGA GGGCGACGAG GAGGAGGAAG GTGGTGTCTG GCGGACTTCC
1701  TACGAGGAGA ACCCGGACTT CGAGGGCATC CCGTGCTGAG AGCTGGTCTG
1751  CTCCATGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGGAGGGCG
1801  GCTGCACATT GGTGAACCTT TTGCAGAGAA CAGAGGAGGA GGAGGACCTG
1851  GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG TGGAGCAGGA
1901  GGTGGGCCCC CCACTGCTAA CGCCACTTTT AGAAGATGCA GAAATCATGC
1951  ACCTGGGACC CTGGACCAAC CGCCTGTCTT GCAGCCTCTG CCGCAGTATC
2001  TCAGTGGCGG TTGTGCGCTC CAACCTCTGG CCGGGGGGCT ATGCTATGCG
2051  CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTACA
2101  GCCCGGAGAG CTTCAACCCG GCCTGCCAGC CCCCATTCA ACAAGAGTAC
2151  CCGAGTGGCC CAGAGATCAT GGAACCAAGC CCTGGGAGCC ACAGAGGAGG
2201  GCAGGCTCTG AAAGCAGCCC AGGAGAGGCG AGGAGACAGA TGACTGAGGC
2251  AGGAGGAGG CGAGGAGGAG GAGGAGGCGG AGGAGACAGA TGACTGAGGC

```

```

2301 CCACCCCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCTCTTA
2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCTCCAGA
2401 GGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAAATAAA TTCCTCCACG
2451 GCATTAAAAA AAAAAAAGAA AAAAAAAG

```

BLAST Results

No BLAST result

Medline entries

86251010:
Molecular cloning and expression of flagellar radial spoke and dynein genes of *Chlamydomona*

81142496:
Radial spokes of *Chlamydomonas* flagella: polypeptide composition and phosphorylation of stalk components.

9450971:
Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm axonemes: involvement of the protein in the regulation of sperm motility.

Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717
Category: strong similarity to known protein

```

1 MGDLPYPYPER PAQPPGRRRT SQASQRRHRS DQAQALAADP EERQIIPDA
51 QRNAPGWSQR GSLSQQENLL MPQVFQAEAA RLGGMETPSV NTGFFSEFQP
101 QPYSDESRMQ VAELTTSMLL QRLQQGQSSL FQQLDPTFQE PPVNPGLQFN
151 LYQTDQFSEG AQHGPIYRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ
201 NAKAYLLQTS INCDLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTQWE
251 WFHPKLDLTL DDPEMQPTYK MAEKQKALFT RSGGGTEGEQ EMEEVGETP
301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KOLVEQQPIH TCRFWGKILG
351 IKRSYLVAEV EFREGEEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV
401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV
451 NARKIKKFFT GYLDTPVVSF PFFPGNEANY LRAQIARISA ATQVSPGLFY
501 QFSEEEGDEE EGGGAGRDSY EENPDFEGIP VLELVDSMAN WVHHTQHILP
551 QGRCTWVNPL QKTEEEEDLG EEEKADGEP EEVEQEVGPP LTPPLSEDAE
601 IMHLAPWTTT LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH
651 KYSPESFNPA LPAPIQQEYP SGPEIMEMSD PTVEEQALK AAQEALGAGT
701 EEEEEGEEEE EGEETDD

```

BLASTP hits

Entry U73123.1 from database TREMBL:
product: "radial spokehead"; *Strongylocentrotus purpuratus* radial spokehead mRNA, complete cds.
Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:
radial spoke protein 6 - *Chlamydomonas reinhardtii*
Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15i5, frame 3

Report for DKFZphtes3_15i5.3

```

[LENGTH] 717
[MM]      80913.61
[pI]      4.36

```

```

(HOMOL)      TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus
radial spokehead mRNA, complete cds. 1e-130
(PROSITE)    TRANSFERRIN_1      1
(PROSITE)    MYRISTYL            5
(PROSITE)    AMIDATION          2
(PROSITE)    CAMP_PHOSPHO_SITE  2
(PROSITE)    CK2_PHOSPHO_SITE   14
(PROSITE)    TYR_PHOSPHO_SITE   1
(PROSITE)    GLYCOSAMINOGLYCAN  1
(PROSITE)    PKC_PHOSPHO_SITE   8
(PROSITE)    ASN_GLYCOSYLATION  1
(KW)         All_Alpha
(KW)         LOW_COMPLEXITY     21.48 %

```

```

SEQ  MGDLPYPYPERPAQQPPGRRTSQASQRRHSRDQAQALAADPEERQQIPPDARQAPGWSQR
SEG  .....XXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccccccc

SEQ  GSLSQQENLLMPQVFAEEARLGGMEYPSVNTGFPSEFQPPQPSYDESRMQVAELTTSML
SEG  .....XXXXXXXXXXXX.....
PRD  cccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhh

SEQ  QRLQOQSLSLFLQLDPTFQEPVPLGQFNLYQTDQFSEGAHQGPYIRDDPALQFLPSEL
SEG  xxxxxxxxxxxxxxxxxxx.....
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPELELAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNQRPEPLSVL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhccccchhh

SEQ  ESLNRTTQWEFHPKLDLTDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEVEGETP
SEG  .....XXXXXXXXXXXX.....
PRD  hhhchhhhhccccccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhcc

SEQ  VPMIMETAFYFEQAGVGLSSDESFRILAMKQLVEQQPIHTCRFWGKILGIRSYLVAEV
SEG  .....
PRD  cccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhccccchhhhh

SEQ  EFREGEEAEVEEEMTEGGEVMEAHGEEEGEEDKEKAVDIVPKSVWKPVPVPIKEESR
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhccccccccccccccccchhhheeecccccccccccccccccc

SEQ  SGANKLYFVFCNEPGLPWTRLPHVTQAIVNARKIKKFTGYLDTPVVSYPPFPNGEANY
SEG  .....
PRD  cccceeeeeccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  LRAQIARISAAQVSLPLGFYQFSEEGDEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  .....XXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccchhhhh

SEQ  WVHHTQHILPQGRCTWVNPLOKTEEEEDLGESEEKADGPEEVEQEVGPPLLTPLSEDAE
SEG  .....XXXXXXXXXXXX.....
PRD  hhhccccccccccccchhhhhhhhhccccchhhhhcccccccccccccccccccccccccc

SEQ  IMHLAPWTRLSCSLCPQYSVAVVRSNLWPGAYAYASGKKFENIYIGWGHKYSPEFNP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LPAPIQEQYPSGPEIMEMSDPTVEEEQALAAQEQALGATEEEEEEEEEEGEETDD
SEG  .....XXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

```

Prosites for DKF2phtes3_15i5.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

WO 01/12659

PCT/IB00/01496

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKFZphtes3_15i5.3)

DKF2phtes3_15j18

group: testes derived

DKF2phtes3_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGGCGGCCCG TGTGCCGCC TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACTCAGGT GGAATGCCCT ACCCCAGTCC TGCGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACACC CTCATCCCTG CACCCCTTCC CTGCCCTGGA
451 TTCTTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGCTCTCTT TGCTTTCCAG CCAGTGTITT GGGGTTTGAA
601 GTTGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAAT GTGTCAGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTTCTCAT
751 CAGGTGATTC TTGGTTGTAG CAAAGTTTAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTG TGAAACTCTA AAAAAAAAAA
851 AAAAAAAAAA GGGCGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPAPWRD VLQQPGVGGE GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHFFSEGV GTQVECLTPV LRLESDMART
101 APHPSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKF2phtes3_15j18, frame 2

Report for DKF2phtes3_15j18.2

```

(LENGTH)      148
(MW)           15665.78
(pI)           8.91
(PROSITE)      MYRISTYL      3
(CK2_PHOSPHO_SITE) 1
(KW)           Irregular

SEQ  MFGCPVRCPKPPTQLISGEASAARLPARVDVLQPGVGEGGLRISWQGAPKSRVRPAFI
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPVPFTVLQSQHYHPFSEGVGTQVECLTPVLRLESMDARTAPHPSSLHPFPAWDSSSPVH
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  CGAPLPSAHGGFPARAEGSWSQPGAGS
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
    
```

Prosites for DKF2phtes3_15j18.2

PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	49->55	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phtes3_15j18.2)

DKF2phtes3_15j3

group: nucleic acid management

DKF2phtes3_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276c, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to be a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits
YGR276c = ribonuclease H
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp

Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CCTTCTTTCG CAGGCAGACG
151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGAAGAGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAGG CAGCCAGGCC
251 CCAAAATAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTAGTGACAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
401 TATGCAGTTC TGGGCAATC CAATGTTCCA AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACCACTAA ACAACGTAGT GGTTTTGTGT CTGCAGGGAA
501 TGAGTCAGCT ACACTTTAC AGGTTCTATT TGGAGTTTGG ATGCTTTCGA
551 AAAGCATTCG GACATAAAT CCGCTTGCTC CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAACCTGAACA AAGAGCTGGA GATCTGCCCA
651 AGACAATGGA AGGGCCCTTA CCTTCTAATG CAAAGCCCGC CATCAACCTT
701 CAGGATGATC CCATCATTCA AAGTATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCCCT CTGACAAAGG AGGAAATGAG AACGTTTAC TTTCATTAC
801 AAGGTTTTCC TGATTGTGAA AACTTTTAC TTACCAAAAT TAATGGTTCT
851 ATAGCAGACA ATAGTCTCTC CTTTGGACTT GACTGTGAAA TGTGCTCTAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGTTGCTT GAAGGAGGCT
951 GCTGTGTTAT GGATGAACCT GTCAAACCTG AAAACAAGAT TCTGACTAC
1001 CTCACCAAGT TTTCCGGGAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTGGG ATCTCAGAGC ACTGAAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAAT TCTTAGCCAA AGTTATTTTG GGAAGGATA
1251 TACAGTGTCG AGACAGACTT GGTGATGATG CCACAGAAGA TGCTAGAAC
1301 ATCCTTGAAT TGGCTCGGTA TTTCTTAAG CATGGCCCAA AAAAGATTGC
1351 AGAATAAAT CTAGAAAGAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTG AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGTCTTTC AAATAAAGAG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGGCTT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACGTGCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTTAAAGCTT TGGCCAGTCC CAGTCAATGA CTTTGTCTCT
1801 TGAAACCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCAGC CTGATTGTG
1851 ACACCTCGT GAATGAGCTG GAAGGAGATT CTGAAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTT CTTGGCCTGG AAGCTGTGAT CTTGCCTAAA GATCTTAAAA
2001 GTGGAAGACA GAAAAATAC TGTTTCTGTA AATTCAAAAG TTTTGGCAGT
2051 GCCCAGCAGG CCTCAACAT TCTCACAGGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCA GGCACCTCCA TGCTGGGCTC AGAGGCTTAC
2151 CACCTGAATC AACAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTGCAGAC TCTGAACTG GACCACCCGA AGATAGCAGC
2251 TGGCGCTGG AGCCGGAAGA TTGAAAGCT CTACAACAGC TTGTGCCCGG
2301 GCACTCTCTG CCTCATCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGGAAATAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCCCTG TGTTCTGTAG TCGGCTGACC ATGTTTCCAT GTGCCATTTC
```

```
2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTACGCTG TAGCCTCCCC
2501 AACCAGCAGA CAGTTTATG GAAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA
2651 AGGGCGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743
Category: similarity to known protein

```
1 MEPEEREGTER HPRKVRERQ APNKLVGAAE AMKAGWDL EE SQPEAKKARL
51 STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVGLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LDCMCLTSK GRELTRISLV
251 AEGGCCVMEDE LVPENKILD YLTSFSGITK KILNPVTTKL KDVRQLKAL
301 LPPDAVLVGH SLDLDRALK MIHPYVIDTS LLYVREQGRR FKLFLAKVI
351 LGKDIQCPDR LGHDATEDAR TLELARYFL KHGPKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLDCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQK YCFLKFKSFG SAQALNILT GKDWKLGRRH ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQALQTLK LDHPKIAAWR WSRKIGKLYN
701 SLCPTGLCLI LLPGKSTHG SLGGLGLMGI KEEESAGPG LCS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15j3, frame 2

TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product";
Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,
N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid
C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces
cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637_9 gene: "SPAC637.09"; product: "putative
exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P
= 2.8e-27

>TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo
sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.
Length = 547

HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
Identities = 358/373 (95%), Positives = 358/373 (95%)

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 164
MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN
Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 60

Query: 165 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD 224
AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD
Sbjct: 61 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD 120

```
Query: 225 NSPLFGLDCM-----CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 269
      NSPLFGLDCM          CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL
Sbjct: 121 NSPLFGLDCMARTTFNFSIGVLQAECLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 180

Query: 270 DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMHPYVIDT 329
      DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMHPYVIDT
Sbjct: 181 DYLTFSFGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMHPYVIDT 240

Query: 330 SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHODATEDARTILELARYFLKHGPKKIAE 389
      SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHODATEDARTILELARYFLKHGPKKIAE
Sbjct: 241 SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHODATEDARTILELARYFLKHGPKKIAE 300

Query: 390 LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDADELPSR 449
      LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDADELPSR
Sbjct: 301 LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDADELPSR 360

Query: 450 NCQTIKCLSNKEV 462
      NCQTIKCLSNKEV
Sbjct: 361 NCQTIKCLSNKEV 373

Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
Identities = 175/179 (97%), Positives = 177/179 (98%)

Query: 538 LETRQVQRPVTELTDCDTLVNELEGDSNQSIIYLSGVSETFKEQLLQEPRLFGLG 597
      L ++VQRPVTELTDCDTLVNELEGDSNQSIIYLSGVSETFKEQLLQEPRLFGLG
Sbjct: 368 LSNKEVQRPVTELTDCDTLVNELEGDSNQSIIYLSGVSETFKEQLLQEPRLFGLG 427

Query: 598 ILPKDLKSGKQKCYCFLKFSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE 657
      ILPKDLKSGKQKCYCFLKFSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE
Sbjct: 428 ILPKDLKSGKQKCYCFLKFSFGSAQQALNLTGKDWKLKGRHALTPRHLHAWLRGLPPE 487

Query: 658 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 716
      STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK
Sbjct: 488 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 546
```

Pedant information for DKFZphtes3_15j3, frame 2

Report for DKFZphtes3_15j3.2

```
{LENGTH} 743
{MW} 83536.58
{pI} 8.87
{HOMOL} TREMBL:AC004381.4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
Chromosome 16 BAC clone CIT987SK-44M2, complete sequence. 0.0
{FUNCAT} 01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
{FUNCAT} 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 1e-10
{FUNCAT} 04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S.
cerevisiae, YGL094c] 1e-10
{FUNCAT} 03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
{PROSITE} MYRISTYL 5
{PROSITE} AMIDATION 1
{PROSITE} CK2_PHOSPHO_SITE 8
{PROSITE} TYR_PHOSPHO_SITE 1
{PROSITE} GLYCOSAMINOGLYCAN 1
{PROSITE} PKC_PHOSPHO_SITE 16
{PROSITE} ASN_GLYCOSYLATION 2
{PFAM} RNA recognition motif. (aka RRM, RBD, or RNP domain)
{KW} Alpha_Beta
```

```
SEQ MEPEREGTERHPRKVRRESRQAPNKLVGAAEAMKAGWDLEESQPEAKKARLSTILFTDNCE
PRD cchhhhhccccchhhhhhhhechhhhhhhhhccccccccchhhhhcccccccccc
```

```
SEQ VTHDQLCELLKYAVLGSNVPKPSWCQLFQHNHLNNVVVFLQGMSQLHFYRFLYLEFGCL
PRD eehhhhhhhhhhhhhccccccccceeeccccccccceeeccccccccchhhhhhhhhhhhhhh
```

```
SEQ RKAFRHKFRLPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSNAKAAINLQDDPIIQKY
PRD hhhhhhhccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhcccccccc
```

```
SEQ GSKKVGLTRCLLTKEEMRTFHFPLQGFDPDCENFLLTKNGSIADNSPLFGLDCMCLTSK
PRD cccccchhhhhhhhhhhhhhhccccccccceeeccccccccceeecccccccccc
```

```
SEQ GRELTRISLVAEGGCCVMDLVKPKENKILDYLTFSFGITKKILNPVTTKLKDVQRQLKAL
PRD cchhhhhheeeccccceeeccccceeeccccccccccccccccchhhhhhhhhhh
```

```

SEQ      LPFDPAVLVGHSLDLDRALKMHPYVITDLSLYVREGRRRKFLKFLKVLIGKDIQCPOR
PRD      eeeccccccccchhhhhhhhhhhccccceeeccccccccchhhhhhhhhhhcccccc

SEQ      LGHDATEDERTILRLARYFLKHGPKCIAELNEALANHOEIOQAQGEFKPNTAEVLQHPNT
PRD      cccccchhhhhhhhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhccccccccceeecccc

SEQ      SVLECEEDSVGQKLLFLTRETDAEGLSPNSRNCQTIKLSNKEVLEQARVEIPLFFSVSIQF
PRD      eeeeeccccccccceeecccccccccccccccccccccccccccccccccccccccccccc

SEQ      SFKAFSPVLTEEMNKRMRIKWTEISTVYAGFPSKNCNLRALKRFLSGFGVQSMTFVLET
PRD      eeeceeehhhhhhhhhhhhhhhhheeeccccccccchhhhhhhhhhhccccceeehhhhh

SEQ      RQVQRPVTELTLDCTLVNELEGDSENGCSYILSGVSETFKEQLHQEPLFLGLGEAVLIP
PRD      cccccccccccccchhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccccceee

SEQ      KDLKSGKKYCKFLKFSFGSAQQALNLTGDKWDLKGRHALTPRHLLHAGWLGPPESTR
PRD      cccccccccceeeccccccccchhhhhhhhhccccccccccccccccchhhhhhhcccccc

SEQ      LPLGRVVPFFFEQALQTLKLDHPKPIAAWRSRKIGKLYNSLPGTCLLLPLGKSTHG
PRD      cccccccccccccchhhhhhhhhchhhhhhhhhhhhhhhheeeccccceeecccccccccc

SEQ      SLSGGLGMLGKEEESAGPGLCS
PRD      cccccccccchhhhhcccccccccc

```

NON GLYCOSYLATION PROCA

PS000001	219->223	ASN_GLYCOSYLATION	PDCC000001
PS000001	419->423	ASN_GLYCOSYLATION	PDCC000001
PS000002	723->727	GLYCOSAMINOGLYCAN	PDCC000002
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	182->185	PKC_PHOSPHO_SITE	PDCC000005
PS000005	238->241	PKC_PHOSPHO_SITE	PDCC000005
PS000005	279->282	PKC_PHOSPHO_SITE	PDCC000005
PS000005	287->290	PKC_PHOSPHO_SITE	PDCC000005
PS000005	447->450	PKC_PHOSPHO_SITE	PDCC000005
PS000005	453->456	PKC_PHOSPHO_SITE	PDCC000005
PS000005	458->461	PKC_PHOSPHO_SITE	PDCC000005
PS000005	481->484	PKC_PHOSPHO_SITE	PDCC000005
PS000005	579->582	PKC_PHOSPHO_SITE	PDCC000005
PS000005	605->608	PKC_PHOSPHO_SITE	PDCC000005
PS000005	630->633	PKC_PHOSPHO_SITE	PDCC000005
PS000005	643->646	PKC_PHOSPHO_SITE	PDCC000005
PS000005	658->661	PKC_PHOSPHO_SITE	PDCC000005
PS000005	678->681	PKC_PHOSPHO_SITE	PDCC000005
PS000005	692->695	PKC_PHOSPHO_SITE	PDCC000005
PS000006	41->45	CK2_PHOSPHO_SITE	PDCC000006
PS000006	193->197	CK2_PHOSPHO_SITE	PDCC000006
PS000006	221->225	CK2_PHOSPHO_SITE	PDCC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDCC000006
PS000006	421->425	CK2_PHOSPHO_SITE	PDCC000006
PS000006	458->462	CK2_PHOSPHO_SITE	PDCC000006
PS000006	579->583	CK2_PHOSPHO_SITE	PDCC000006
PS000006	630->634	CK2_PHOSPHO_SITE	PDCC000006
PS000007	370->379	TYR_PHOSPHO_SITE	PDCC000007
PS000008	27->33	MYRISTYL	PDCC000008
PS000008	186->192	MYRISTYL	PDCC000008
PS000008	575->581	MYRISTYL	PDCC000008
PS000008	714->720	MYRISTYL	PDCC000008
PS000008	720->726	MYRISTYL	PDCC000008
PS000009	337->341	AMIDATION	PDCC000009

1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 26

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)	
HMM	*IYVGNLPWDtTEEDLRdLfGQFGpIvsIRmMRdReTGRSGrGfAFvEFED	
	IY+ +++ +T +E+L + + F + + + +D G+ + ++F +F++	
Query	571 IYtLGSVt-EtFtLQGLvPRfLFGLEAvILPKDLKSGRQKkyCYfLKfKS	618
HMM	EEDaekAIdemNG..meFmGrRrV*	
	+A+ A+ + G ++ GR +	
Query	619 FGSAQqALNlLTGkDkWlKGRHALT	643

DKF2phtes3_15k11

group: signal transduction

DKF2phtes3_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCGCGCGTG CCAACCCCTCC
51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGGT CCGGGTGGGG
151 TTCTACGACA TCGAGGGCAC GCTGGGCAAG GGCAACTTCG CTGTGGTGAA
201 GCTGGGGCGG CACCGSATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAACTTA CCGAGAAGTA
301 CAAATAATGA AAATGTTAGA CCACCCCTAC ATAATCAAAC TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATCTGGGCA AATCCTGTCT GCTGTTGATT ATTGTCATGG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAACTTC CTGCTGGATA
551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGGAAA TTTCTTTAAA
601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCTTT ATGCAGCCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAGG ACCACAGCTG GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCCCTGTCT GTGGAGCTCT GCCCTTTGAT
751 GGACCGGACT TTCCAATTTT GAGGCGAGGG GTTCTGGAAG GAAGATTCCG
801 GATTCCGTAT TTCTATGTCA AAGATTGCGA GCACCTTATC CGAAGGATGT
851 TGGTCCTAGA CCCATCCAAA CGGCTAACCA TAGCCCAAT CAAGGAGCAT
901 AAATGGATGC TCATAGAAGT TCCTGTCCAG AGACCTGTTT TCTATCCACA
951 AGAGCAAGAA AATGAGCCAT CCATCGGGGA GTTTAATGAG CAGGTTCCTGC
1001 GACTGATGCA CAGCCTTGGG ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1051 CAGAACAAGA GCTATAACCA CTTTGTCTGC ATTTATTTCT TGTGGTGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTTGATG
1151 GCGCCGACGG TCGGCCTAGC ACCATTGCTG AGCAACAGT TGCCAAGGCA
1201 CAGACTGTGG GGCTCCCAGT GACCATGCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCTCCCCCC AGGCATCCAA CGTGGAGGCC TTTTCATTTC
1301 CAGCATCTGG CTGTGAGGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCTGTGTC CTCTGTCTCT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCCAG CAACATGATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCCTCAT
1501 GCCTTTGAGG CATTTCAGTC CACACGCAGC GGGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCGTGAT GCCTGGGGCA GGGAAAAATT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGAGTATGAT
1651 ATGGGGTCTG TTCAGAGGGA CCTGAACTTT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCAGCCA TGCAAGGCTCT GAGCTCCCG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGGCCGCG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATT AGACAACTC
1901 TTCAGAATCT GGCTAGAACC AAAGGAATTC TAGAGTTGAA CAAAGTGCAG
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCCTAACC TGGCGCCGCG
2001 GGCTCCTCAG CTCAGGAGCC TTGCTAGCAG CTGCCCTCAG GAAGAAGTTT
2051 CTCAGCAGCA GGAAGCGCTC TCCACTCTCC CTGCCCGCTG GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCGAGC CTTCTGTCAA AGGCCAGAA CACCTGTGAG CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACTGCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCCCCTCCCG CCAGGAGACT CCACCGCCTT CTCAGCAGG CCCACCGTTC
2401 AGCCTGACCC AGCCCTGAG CCCCCTGCTG GAGCCTTCTT CCGAGCAGAT
2451 GCAATACAGC CTTTCTCTCA GCCAGTACCA AGAGATGCAG CTTAGCCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGCTC TCCTCTGCCC CACGCAAGTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCACC CCTCCACCAC CACGACAGCC
2601 AGGAGCTGCC CCAGCCCCCT TACAGTTCTC CTATCAGACT TGTGAGCTGC
```

```
2651 CAAGCGCTGC TTCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCAGAGAG
2751 CCCAGGACTG CAAGAGGGCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACGCTGT GGATCCACAA
2851 CACAACGGGT ATGTCTTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTTCTAT TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGCT GGATGTTGCT TCCTCTGCTG
3051 TCTGCCCCAC CACAAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTGAGTGGAA GCAAGCTCTC GAGGCGAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATATCAAA GGGCAACCTT GGTGAAAGCA GAAAGGCTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCAGTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCTATT AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCGAG GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCTGCGAGT TTCTCGTGGG CACTGTGGCG TTGCGGGCAG TCGGTTCCCA
3751 GGGTACCTGT TGTCTCTTTT CCGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTGGAC AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAGG ACATGATAAA TGAACCTTTT TGTCCTATGT
3901 GAAGTGGTAG TGGGTGGCTT TTCCCCCAGA TCATGCTTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTTATG TCAATGCTAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTCCCA CCAATGTTGA ATCAGCATTC TTAATTTCTG
4051 TAAAGTTTGG ACTTGTAAAG AAATGTTCAA GTATTACAGC AATATTCAAA
4101 GAAAGAACCA CAGATGTGTT AACCATTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTAA ACCAACACTT ACAATTCAGT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTGGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGT
4301 AATTCTAAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAACACCT TTAAGTTGCC TTTCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGGTTCCATT
4501 CTGGCCATGC GAGCCGAGCT CTAACCAACG TCGGTAACTT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTGCCCA GCGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACTG TGCTCACGGT TGGACTTGGT GTGAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAAATA AAGAACATCA TTTCAATTTA
4801 AAAAAAAGG AAAAAAGG CCGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTAAAAA AAAAAAG
```

BLAST Results

Entry HSG4921 from database EMBL:
human STS SHGC-37164.
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:
Homo sapiens mRNA for KIAA0781 protein, partial cds.
Score = 10725, P = 0.0e+00, identities = 2145/2145

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959
Category: known protein

```
1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQGPVVRVG
51 FYDIEGLTGG GNFVVKLGR HRITKTEVAI KIIDKSOLDA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVNET KSMILYLVTEY AKNGEIPDYL ANHGRRLNESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPFVF EGQQYEGPQL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSDECHELI RRLMLVDPFK RLITIAQIKH
```



```

301 KWMLEIEVPVQ RPYLYPQEQE NEPSIGEFNE QVRLMHSLG IDQKTIESTL
351 QNKSYNHFAA IYFLVVERLK SHRSSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTHM SPNMRLLRSA LLPQASNVEA FSFPASGCQA EAAFMEEECV
451 DTPKVNGLCL DPVFPVLVRK GCQSLPSNM ETSIDEGLET EGEAEEDPAH
501 AFEAFQSTRS QRRRTLSEV TNQLVVMFGA GKIFSMNDSP SLDSDVSEYD
551 MGSVQDLNLF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVNHRSPV SFREGRRASD TSLTQGI VAF RQHLQNLART KGILELNKVPQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLETQ YLQHLRQKPS LLSKAQNTCQ LYCKEPPRSL EQQLQEHRLQ
751 QKRLFLQKQS QLQAYFNQMQ IAESSYPQPS QQLPLPRQET PPPSQAPPFF
801 SLTQPLSPVL EPSSEQMOMYS PFLSQYQEMQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLQFSYQT CELPSAASPA PDYPTPCQYP
901 VDAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPLGLFD CEMLDAVDQ
951 HNGYVLVN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15k11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15k11, frame 1

Report for DKFZphtes3_15k11.1

```

[LENGTH]      926
[MW]           103915.77
[pI]           5.70
[HOMOL]        TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT]       11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT]       03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT]       03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT]       02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT]       01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT]       03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT]       10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT]       10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT]       04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT]       10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT]       10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT]       04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT]       05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT]       01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]
 2e-14
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14
 [FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YBR097w] 1e-10
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]
 1e-10
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]
 1e-10
 [FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]
 4e-09
 [FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.
 cerevisiae, YHR079c] 1e-07
 [FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c]
 1e-07
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04
 [BLOCKS] BL00415A Synapsins proteins
 [BLOCKS] BL00239B Receptor tyrosine kinase class II proteins
 [BLOCKS] BL00107A Protein kinases ATP-binding region proteins
 [SCOP] dlgo1_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-78
 [SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 1e-81
 [SCOP] dlkoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 5e-89
 [SCOP] dlkoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 5e-86
 [SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 3e-80
 [SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 6e-70
 [SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 1e-95
 [SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom)] 7e-71
 [SCOP] dlydse_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 2e-96
 [SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom)] 2e-72
 [SCOP] dlcdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 5e-97
 [SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma] 2e-68
 [SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 3e-53
 [SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 3e-78
 [SCOP] dickia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 1e-58
 [EC] 2.7.1.117 Myosin-light-chain kinase 3e-49
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 4e-78
 [EC] 2.7.1.38 Phosphorylase kinase 3e-41
 [EC] 2.7.1.37 Protein kinase 7e-45
 [EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42
 [EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78
 [PIRKEW] phosphotransferase 3e-93
 [PIRKEW] nucleus 2e-74
 [PIRKEW] calcium 2e-40
 [PIRKEW] transferase 3e-33
 [PIRKEW] duplication 2e-32
 [PIRKEW] tandem repeat 7e-45
 [PIRKEW] phorbol ester binding 4e-33
 [PIRKEW] zinc 4e-33
 [PIRKEW] ion transport 1e-32
 [PIRKEW] cell cycle control 1e-45
 [PIRKEW] serine/threonine-specific protein kinase 2e-97
 [PIRKEW] oncogene 1e-34
 [PIRKEW] phospholipid binding 2e-32
 [PIRKEW] autophosphorylation 2e-74
 [PIRKEW] brain 6e-36
 [PIRKEW] heterotetramer 8e-38
 [PIRKEW] mitosis 1e-45
 [PIRKEW] polymer 5e-41
 [PIRKEW] magnesium 6e-80
 [PIRKEW] ATP 2e-97
 [PIRKEW] polyprotein 1e-34
 [PIRKEW] alternative initiators 2e-31
 [PIRKEW] phosphoprotein 2e-74
 [PIRKEW] apoptosis 8e-38
 [PIRKEW] cGMP binding 4e-33
 [PIRKEW] glycoprotein 3e-36
 [PIRKEW] skeletal muscle 8e-38
 [PIRKEW] protein kinase 2e-50
 [PIRKEW] testis 5e-41
 [PIRKEW] cAMP binding 8e-38
 [PIRKEW] transforming protein 4e-33
 [PIRKEW] purine nucleotide binding 7e-52
 [PIRKEW] calcium binding 7e-45
 [PIRKEW] alternative splicing 5e-42
 [PIRKEW] P-loop 7e-52
 [PIRKEW] lipoprotein 8e-38
 [PIRKEW] proto-oncogene 4e-33
 [PIRKEW] segmentation 1e-34
 [PIRKEW] core protein 1e-34

```

(PIRKW]      muscle 8e-38
(PIRKW]      myristylation 8e-38
(PIRKW]      EF hand 7e-45
(PIRKW]      cell division 3e-49
(PIRKW]      homodimer 1e-32
(PIRKW]      calmodulin binding 5e-42
(SUPFAM]     ribosomal protein S6 kinase II 1e-34
(SUPFAM]     calcium-dependent protein kinase 7e-45
(SUPFAM]     AMP-activated protein kinase 6e-80
(SUPFAM]     protein kinase akt 3e-36
(SUPFAM]     protein kinase SPK1 7e-41
(SUPFAM]     unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase 5e-42
(SUPFAM]     calmodulin repeat homology 7e-45
(SUPFAM]     cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
(SUPFAM]     protein kinase DUN1 6e-36
(SUPFAM]     protein kinase C zeta 4e-33
(SUPFAM]     Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
(SUPFAM]     death-associated protein kinase 8e-38
(SUPFAM]     pleckstrin repeat homology 3e-36
(SUPFAM]     ankyrin repeat homology 8e-38
(SUPFAM]     protein kinase homology 8e-99
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase II 6e-38
(SUPFAM]     protein kinase C zinc-binding repeat homology 4e-33
(SUPFAM]     protein kinase C delta 2e-32
(SUPFAM]     cGMP-dependent protein kinase 3e-33
(SUPFAM]     protein kinase cdrl 1e-45
(SUPFAM]     kinase-related transforming protein 2e-50
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase I 8e-42
(SUPFAM]     kinase interaction domain homology 7e-41
(SUPFAM]     gag-akt polyprotein 1e-34
(PROSITE]    PROTEIN_KINASE_ATP      1
(PROSITE]    MYRISTYL                 3
(PROSITE]    AMIDATION                 2
(PROSITE]    CAMP_PHOSPHO_SITE        4
(PROSITE]    CK2_PHOSPHO_SITE         15
(PROSITE]    TYR_PHOSPHO_SITE          2
(PROSITE]    PKC_PHOSPHO_SITE          10
(PROSITE]    ASN_GLYCOSYLATION         2
(PROSITE]    PROTEIN_KINASE_ST        1
(PFAM]       Eukaryotic protein kinase domain
(KW]         Irregular
(KW]         3D
(KW]         LOW_COMPLEXITY      12.31 %

```

```

SEQ      MVMADGPRHLQRGPPVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVAVN
SEG      .....
1ctpe    .....EEECTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ      LEKIYREVQIMKMLDHPHIKLYQVMEKSMPLYLVTEYAKNGEIFYDLANHGRLNESEAR
SEG      .....
1ctpe    HHHHHHHHHHHHCCCTTTBCCBBBBBETTEEEEEECTTTTBHHHHHHHHHCCCHHHHH

SEQ      RKFQILSAVDYCHGRKIVHRDLKAENLLDNMNIKIADFGGNFFKSGELLATWCGSP
SEG      .....
1ctpe    HHHHHHHHHHHHHHCCCECCCGGGEEETTTTCEECCTTTTEETT-TTBC-CCCCCG

SEQ      PYAAPEVFEGQQYEGPQLDIWSMGVVLYVLVCGALPFDGPTLPILRQVLEGRFRIPIYFM
SEG      .....
1ctpe    GGCCHHHHHHCCBC-HHHHHHHHHHHHHHHHCCCTTTTTHHHHHHHHHHCCCTTTT

SEQ      SEDCEHLIRMLVLDPSKRLTIAQIKENKMWLIEVPVQRPVLYPQEQENEPSIGEFNEQV
SEG      .....
1ctpe    CHHHHHHHHHTTTTGGGTTTHHHHHHCGG.....

SEQ      LRLMHSLGIDQQKTIESLQNSYNHFAAIYFLLVERLKSRRSSFPVEQRLDGRQRRPSTI
SEG      .....
1ctpe    .....

SEQ      AEQTVAKAQTVGLPVTMHSPNMRLRLSALLPQASNVEAFSPASGCQAEAAFMEEECVDT
SEG      .....
1ctpe    .....

SEQ      PKVNGCLDPVPPVLRKGCQSLPSNMETSIDEGLETEGEAEEDPAHAFAEAFQSTRSGQ
SEG      .....
1ctpe    .....XXXXXXXXXX.....

SEQ      RRHTLSEVTNQLVMPGAGKIFSMNDSPLSDVDSEYDMGSVQRDNLFLDNPSLKDIML
SEG      .....
1ctpe    .....

```

```

SEQ      ANQPSFRMTSPFISLRPTNPAMQALSQKREVNRSFVSFREGRRASDTSLTQGI VAFRQ
SEG      .....
lctpE    .....

SEQ      HLQNLARTKGILELNKVQLLYEQIGPEADPNLAPAPQLQDLASSCPQEEVSQQQESVST
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpE    .....

SEQ      LPASVHPQLSPRQSLETQYLQHRLQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQOK
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpE    .....

SEQ      RFLQKQSQLQAYFNQMQIAESSYPQPSQQLPLPRQETPPPSQQA PPSLTQPLSPVLEP
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpE    .....

SEQ      SSEQMYS PFLSQYQEMQLQLPSTSGPRAAPLPTQLQQQPPPPPPPPPPPPPPGGAAPA
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpE    .....

SEQ      PLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQSDLTGPDGFRSPGLQEA PPSYDPLAL
SEG      xxx.....
lctpE    .....

SEQ      SELPGLFDCEMLDAVDPOHNGYVLVN
SEG      .....
lctpE    .....

```

Prosites for DKF2phtes3_15kl1.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKF2phtes3_15kl1.1

HMM_NAME Eukaryotic protein kinase domain

HMM		*YeigRiIGeGsFGtVYKCiWr.TGeIVAIIkkzms.....FlREI	
Query	20	Y I+++G+G+F++V++++R T +VAIKII+K++++ + RE+	68
		YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVNLKIIYREV	
HMM		qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw	
Query	69	QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E	117
		QIMKMLDHPHIIKLYQVME-TKSMLYLVTEYAKNGEIFDYLANHGRLNES	
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIdNgqIKIcDFGLARqM	
Query	118	E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+ ++	167
		EARRKFWQILSAVDYCHGRKIVHRDLKAENLLDNNMNIKIADFGFGNFF	
HMM		nnYerMtrfCGTPWYMMAPEVImg.nyYttkVDMWSFGCILMEMMTGep	
Query	168	+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G +	215
		KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLDIWSMGVLYVLVCGAL	
HMM		PFyddnMemImrIiqrfrpfWpnCSeElyDFMrwCwnyDPeKRPTFrQI	
Query	216	PF++ ++ + + +++ R++++ +SE++ ++R+++ +DP+KR+T+ QI	265
		PFDGPTLPILRQRVLEGRFRIPYFMSDCEHLIRRLVLDPKRLTIAQI	
HMM		LnHPWF*	
Query	266	+H W+ KEHKWM 271	

DKF2phtes3_17f10

group: testes derived

DKF2phtes3_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1 CTTCAAGTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51 TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTCCT AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCG TGAAACATAA ATCGTCGGGA
201 AAGATTTTGT CTAGTGAACA CCCTGAATTT CAACCAGCAA CAACACAGCAA
251 TGAAGAAATTT GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
301 CTAAAAAAGG TCCCCCAGTA CTTTGAAGAG ATGAGCTTAG GGAAGAAAGTA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
401 TGAATATAGAG CCTCCATCAA CAGAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCCAGCC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGCAGAGA ATTCTGTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTTCTGTCTG AAATTCAGCC TCCATCAGCT
651 GAAGAGTCTC CTCTGTAGA GCTTCTGGCT GAAATTCGTC CTCCATCAGC
701 TGAAGAGTCC CCTTCAGAAG AGCCTCCTCG TGAATTCCTG CCTCCACCAG
751 CTGAAAAATG TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTGA AGTACAGCCT TTACCAGCTG AGGGCGCCCT
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCCCTG
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCCTAGAGA AGAGGCTCGA
951 GAACTTCAGC TTTCAACAGC TATGGAGACC CCTGCAGAAG AGGCTCCTAC
1001 TGAATTTTCA TCTCCATTAC CTAAGAGAGC CACTGCAGAA GAGGCTCTGT
1051 CTGAAATTTA GCTTCTAGCA GCTACGGAGC CTCCTGCAGA TGAAATCTCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCCTGCA GAAGAGGCTC
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTCG AGATGAAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCCTTGTA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTCAGCT TCTAGCAGCT ACAGAGGCTT CTGCAGAAGA
1451 GGCTCCTGCT GAAGTTGAGC CTCACCCAGC TGAGGAGGCC CCGCTGAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCTGTA AGTTCAGCCT CCACCAGCTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGGCCCCGCT GAAGTTGAGC
1651 CTCACCCAGC TGAGGAGGCC CCCCCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCTCG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCTATAGA
1751 AGAGACCCCT GCTGAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCCT CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTAC CTCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGTT AGAAGATGTG TCTACAGAAT
1951 TTCAGTCACC GCAGGTGACA GGAATTCAG CAGTAAATTT AGGATCGGTT
2001 GTTTTGGAAG GTGAAGCAA ATTTGAAGAG GTTTCAAAAA TCAATCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAATTAAGAC AACGCTCCTC TGAAGTGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGTTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGG ACCCAAGAC TTGGAATACA GGTGGAATAA TGAACATAAA
2301 AAAGTGTAGC AGCATAAAAT TACTGTGTTT AATTTCATTC AAAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAATTGCT
2401 TAAATATCC TTTCTACAGT AAACCTGTGT ACACGAGTAA AGTTTAATCT
2451 GCAGCCATCT TTTCTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710
Category: similarity to known protein
Classification: unclassified

```
1 MDRSQQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP
101 VLLEDELREE VTPVVVQEGS AVKKVASAEI EPPSTKFPFA KIQPPPLVEEA
151 TAKAEPRPAE ETHVQVQPST EETPDAAEAT AVAENSVKVQ PPPAEAPLV
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPPEAI LPPPAEKSPS
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL
301 LPEEAPREEA RELQLSTAME TPAAEAPTEF QSPLPKETTA EEASAEIQLL
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAEETT EEASAEIQL
401 LAIEAPADET TPAEAQSPLS EETSAAEAPA EVQSPSAGKV SIEEAPLELQ
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA
501 EEAPEAVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE
551 APSEVQPPPA EEAPEAVQSL PAETPIEET LAAVHSPPAD DVPAAEASVD
601 KHSPPADLLL TEEFPIGEAS AEVSPPPSEQ TPEDEALVEN VSTEFQSPQV
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTLEIESVFH
701 IELKQRPEEL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P
= 7.4e-43

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N
= 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat
Length = 1,072

HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43
Identities = 185/622 (29%), Positives = 320/622 (51%)

```
Query: 33 SESEIVVISRPDSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92
SE +I V+ + + + +E + + + + + E E Q E G + + TS
Sbjct: 436 SEEKIKVVEKSEKETVIVEEQTEETQVTEEVTEEDKEAQGEEEEAEEGGEEAATTSPP 495

Query: 93 QETKKGPPVLLEDELREEVTPVVQEGSAVKKVASAEIEPPSTKFPFAKIQPPPLVEEATA 152
E P + +EE P + A K + AE + P+ K PA+++ P ++ A
Sbjct: 496 AEEAASPEKETKSPVKEEAKSPAEAKSPAEAK-SPAEAKSPAENVKSPAENVKSPAENKSPA 554

Query: 153 KAEPRAEETHVQVQPSTETPDAAEATAVAENSVKVQPPPAEEAP-LVEFPAEIQPPSA 211
+A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+
Sbjct: 555 EAKS-PAE---VK-SPATVKSPAENKSPAENKSPAENVKSPATVKSPGEAKSPAENKSPA 609

Query: 212 EESP-SVELLAELPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268
+SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+
Sbjct: 610 VKSPVEAKSPAENKSPASVKSPGEAKSPAENKSPAENVKSPATVKSPVEAKSPAENVKSPVT 669

Query: 269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAPREEARELQLSTAMETPAE-EAP 327
V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P
Sbjct: 670 VKS-PAEA---KSPVEVKSPASVKSPSEAKSPAGAKSPA-EAKS---PVVAKSPAENKSP 721

Query: 328 TEFQSPFLPKETTAEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHAEVQS--- 383
E + P ++ AE S A + PA+ ++PAE+SP+ E S E+A + V+
Sbjct: 722 AEAKPPAEAKSPAENKSP-----AEAKSPAENKSPAENKSPV-EVKSPEKAKSPVKEGAK 775

Query: 384 PLAEETTAEEASAEIQLLAIEAPAD-ETPAEAQSPLEET-SAEAPAE-EVQSPSAGKV 440
```

LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK
 Sbjct: 776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTP 833

Query: 441 SIEEA--PLELQPPSGEETTA--EASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494
 + EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +
 Sbjct: 834 AKEEAKRPADIRSPQVKSPEAKEAKSPEKEETRTKVPAPKKEEVKSPVEEVKAKEPPKK 893

Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553
 V+ P EV+ +EAP E Q P AEE + P +++P E + EEA
 Sbjct: 894 VEEETPATPKTEVKESKKDEAPKEAQKPKAEKEPLETEKP--KDSPEAKK--EEAKE 948

Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAEEASVD-KHS 603
 + P EE PA++ ' ++ P AE+ +E + P ++VPA D K
 Sbjct: 949 KKAAPPEETPAKLGVKKEAKPKAEADAKAKEPSKPSKEKEPKKEEVPAPEKKDTEKE 1008

Query: 604 PPADLLTTEFFIGEASAEVSP--PSEQT-PEDEALVENSTEFQSPQ 649
 + EE P +A A+ P E + P+ E ++ ST+ + Q
 Sbjct: 1009 KTTESKKPEEKPMQAKAKEEDKGLPQEPSPKTEKAESSTSDQKDSQ 1057

Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42
 Identities = 184/628 (29%), Positives = 310/628 (49%)

Query: 18 IPPVEKVDKEQTYFSESEIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74
 I VEK +KE ++E + ++ + E+ + + G+ A+ P + A
 Sbjct: 440 IKVVEKSEKETVIVEEQTEEQTEEVTEEDKEAQGESEEEAEEGGEEAATSPPAEEA 499

Query: 75 TNSNEEIQKNISRTSFTQETKKGPPVLLLEDELREEVTVPVVQEGSAVKKVASAEIEPPS 134
 + +E + + + + K P E + E P + A K + AE + P+
 Sbjct: 500 ASPEKET-KSPVKEEAKSPAEEAKSPA---EAKSPAEEAKSPAEEVKSPEAEVKS-SPAEEAKSPA 554

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQ-PSTEETPDAAEATAVAENSVKVPQPP 193
 K PA+++ P ++ A+A+ ++ +V+ P+T +P+ + A A++ +V+ P
 Sbjct: 555 EAKSPAEEVKSPEATVKSPEAKSPAEEAKSPAEEVKSPEATVKSPEAKSPAEEVKSPEATVKSPEAKSPA 614

Query: 194 AEEAPL-VEFPAEIQPPSAEESPS-VELLAEILPPSAEESPS-EPPAEILPPPAEKSPS 250
 ++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+
 Sbjct: 615 EAKSPAEEVKSPEATVKSPEAKSPAEEAKSPAEEVKSPEATVKSPEAKSPAEEVKSPEATVKSPEAKSPA 674

Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPPEEAPR 307
 + E+SP++ K+P E + P A+ E ++P + P ++ AE + P ++P
 Sbjct: 675 EAKSPVEVKSPEATVKSPEAKSPAEEAKSPAEEVKSPEATVKSPEAKSPAEEVKSPEATVKSPEAKSPA 734

Query: 308 EEARELQQLSTAME--TPAE-EAPTEFQSP----LP-KE---TTAEASAEIQLLAATE-- 354
 E + + E +PAE ++P E +SP P KE + AE S E E
 Sbjct: 735 EAKSPAEEVKSPEATVKSPEAKSPAEEAKSPAEEVKSPEATVKSPEAKSPAEEVKSPEATVKSPEAKSPA 794

Query: 355 -PPAD-ETPAEARSPLSEET-SAEAAH-EVQSPLEAETTAEAS--AEIQLLAATEAPA 408
 PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ ++P+
 Sbjct: 795 KPPAEVKSPEAKSPMKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPQVKSPEAKSPA 854

Query: 409 DETPAEQSPLSEETSAAE-APA--EVQSPSAKGVSEIAPLELQPPSGEETTAEEASAA 465
 E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A
 Sbjct: 855 KE---EAKSPEKEETRTKVPAPKKEEVKSP-----VEEVKAK-EPPKKVE---EETPA 901

Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525
 E+ +EAP E Q P AEE + P +++P E + A+E A P E
 Sbjct: 902 TPKEVKSKEKDEAPKEAQKPKAEKEPLETEKP--KDSPEAKKEEAKKAAA---PEE 956

Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581
 E PA++ + P E+A P++ PSE + P EE PA + +E E+
 Sbjct: 957 ETPAKLGVKKEAKPKAEADAKAKEPSK--PSEKEPKKEEVPAPEKKDTEETTESK 1014

Query: 582 AAVHSPADDVPAEEASVDKHSPPADLL-LTEFFIGEASAEVSPPPSEQTPDEEA 636
 P EE DK P TE+ ++ + PSE+ PED+A
 Sbjct: 1015 KPEEKPMQAKAKEE---DKGLPQEPSPKTEKAESSTSDQKDSQSEKAPEDKA 1067

Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36
 Identities = 162/540 (30%), Positives = 275/540 (50%)

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQSTEETPDAAEATAVAENSVKV 189
 TE P KI P + K+E + +E+ V V+ TEE E T E +
 Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEEQTEEVTE--EEDKEA 474

Query: 190 QPPPAEEAPLVEFPAEIQPPSAEESPSVELLAEILPPSAEE--SPSE-EPPAEILPPPAE 246
 Q EEA A P AEE+ S E E P EE SP+E + PAE P
 Sbjct: 475 QGEEEEEAEEGGEEAATSPPAEEAASPE--KETKSPVKEEAKSPAEEAKSPAEEAKSPA 532

Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPPEEAP 306
 KSP+ E++SP+ K+P E + PAE ++P+V+ P ++ AE + +P
 Sbjct: 533 KSPA-----EVKSPEVKSPEAKSPAEEAKSPAEEVKSPEATVKSPEAKSPAEEVKSPEATVKSPEAKSPA 583

Query: 307 REEARELQLSTAME--TPAE-KAPTEFQSPLPKETTAEAS-AEIQLLAATEPPAD-ETP 361
 E + + E +PAE ++P E +SP+ ++ AE S A + + + PA+ ++P
 Sbjct: 584 AEVKS PATVKSPGEAKSPAEAKSPAEVKS PVEAKSPAEAKSPASVKS PGEAKSPAEAKSP 643

Query: 362 AEARSPLSEETSAB-EAHAEVQSPLAEETTAEASAEIQLLAIEAPAD-ETPAEAQSPL 419
 AE +SP + + + E ++ AEV+SP+ ++ AE A + + + ++PA ++P+EA+SP
 Sbjct: 644 AEVKS PATVKSPVEAKSPAEVKS PVTVKSPAE-AKSPVE---VKSPASVKS PSEAKSP- 697

Query: 420 SEETSAAEAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAE 478
 + ++PAE +SP AK + + ++P E +PP+ ++ AE S A + A + A +
 Sbjct: 698 ----AGAKSPAEAKSPVVAKSPAEAKSPAEAKFPPEAKSPAEAKSPAE---AKSPAEAK- 749

Query: 479 APAEVQPPPAEAPAEVQPPPAEAP--AEVQPPPAEAP--EVQPPPAEAPAEVQPP 534
 +PAE + P ++P + P E A AE + P ++P E ++PP ++P + + P
 Sbjct: 750 SPAEAKSPVEVKSPEKAKSPVKEGAKSLAEAKSPKAKSPVKEIKPPAEVKSPEKAKSP 809

Query: 535 PAEEAPAEVQPPPAEAPSEVQPPPAEEA--PAEVQSLPAEETPIEETLAHVSPPADV 592
 EEA + + + E + P EEA PA+++S ++P +E SP ++
 Sbjct: 810 MKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPQVKSPEKE---AKSPEKEET 866

Query: 593 PAEEASVDKHS--PPADLLTEFFIGEASAEVSPPPSEQTPEDEALVENVSTEFOSQOV 650
 E+ + K P + + +E P + + +E P + + +E Q P +
 Sbjct: 867 RTEKVAPKKEEVKS PVEVKAKEFP--KKVEEEKTPATPKTEVKS KKEAPKEAQKPKA 924

Query: 651 AGIPAVKLGSVVLEGEAKFEVSK 674
 + GEAK EE +
 Sbjct: 925 EEKEPLTEKPKDSPGEAKKEEAK 948

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34
 Identities = 123/390 (31%), Positives = 213/390 (54%)

Query: 308 EEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPA---EA 364
 E+ E+Q++ E EE E Q +E AEE E A T PPA+E + E
 Sbjct: 455 EQTEEIQVT---EEVTEEDKEAQGE---EEEEAEEGGEEA---ATTSPPEEAASPEKET 506

Query: 365 RSPLSEETSAAEHAHVQSPLAEETTAEAS-AEIQLLAIEAPAD-ETPAEQSPLE 422
 +SP+ EE + AE +SP ++ AE S AE++ A +++PA+ ++PAEA+SP +
 Sbjct: 507 KSPVKEAKSP---AEAKSPAEAKSPAEAKSPAEVKS PVEAKSPAEAKSPAEAKSPAEVK 563

Query: 423 TSAE-EAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAEAP 480
 + A ++PAE +SP+ AK + ++P ++P GE EA + + + EA ++P
 Sbjct: 564 SPATVKSPAEAKSPAEAKSPAEVKS PATVKSP-GEAKSPAEAKSPAEVKS PVEA---KSP 619

Query: 481 AEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAP 540
 AE + P + ++P E + P ++PAEV+ P ++P E + P ++P V+ P ++P
 Sbjct: 620 AEAKSPASVKS PGEAKSPAEAKSPAEVKS PATVKSPVEAKSPAEVKS PVTVKSPAEAKSP 679

Query: 541 AEVQPPPAEAPSEVQPPPAEAPAEVQSLPAEETPIEETLAHVSPPAD-DVPAEEASV 599
 EV+ P + ++PSE + P ++PAE +S ++P E A PPA+ PAE S
 Sbjct: 680 VEVKS PAVKS PSEAKSPAGAKSPAEAKSPVVAKSPAEAKSPAEAKFPPEAKSPAEAKSP 739

Query: 600 DKHSPPADLLTEFFIGEASAEVSPPPSEQTPEDEALVENVSTEFOSQOVAGIPAVKLG 659
 + PA+ E ++ EV P ++P E + + + +E +SP+ A P VK
 Sbjct: 740 AEAKSPAEAKSPAE---AKSPVEVKSPEKAKSPVKEG-AKSLA-EAKSPAEAKSP-VK-E 792

Query: 660 SVVLEGEAKFEVSKINSVLKDLSTNDGQAPTLEIES 697
 + E K E +K S +K+ + + +A TL+++S
 Sbjct: 793 EIKPPAEVKSPEKAK--SPMKEEAKSPE-KAKTL DVKS 827

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18
 Identities = 124/420 (29%), Positives = 199/420 (47%)

Query: 252 ELLGEIRSPSAQKAPIEVQPLPA-----EGALEEAPAKVPPPTVEETLAEVQPLLPEEAP 306
 ELLG+I+ A +A + + A AL E A++E TV+ TL +
 Sbjct: 236 ELLGQIQGCGAAQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEENFVRVLD 295

Query: 307 REEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPAEARS 366
 EA ++ + AM + EE TE++ L TT E++ L +T+ + +E
 Sbjct: 296 LSEAAKVN-TDAMSAQEETI-TEYRRLQARTT-----ELEALKSTKESLERQRSELED 347

Query: 367 PLSEE-TSAEEAHAHVQSPLAEETTAEASA--EIQLLAIEAPAD-ETPAEAQSPLSEE 422
 + S ++A ++ + L T E A+ E Q L ++ D E A + EE
 Sbjct: 348 RHQVDMASYQDAIQQLDNEL--NTKWEAAQLREYQDLLNVKMLDIEIAAYRKLEEGE 406

Query: 423 TSAEEAPAEV-----QSPS-AGVSIIE-EAPLELQPPSGEETT-AEEASAAIQLLA-A 471
 P+ + PS + + + +E +++ S +ET EE + IQ+
 Sbjct: 407 CRIGFGPSFSLTEGLPKIPSMSTHIKVKSEEKINVVERSEKETVIVEEQTEIEQVTEEV 466

Query: 472 TEASAEAPAEVQPPPAEAPAEVQPP--PAEEAPAEVQPPPAEAPAEVQPPPAEAPAEVQPP 524
 TE +EA E + AEE E PPAEEA+ E + P EEA PAE + P
 Sbjct: 467 TEEEDKEAQGE-EEEEAEEGGEEAATTSPPAEEAASPEKETKSPVKEAKSPAEAKSPAE 525

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18
Identities = 115/364 (31%), Positives = 166/364 (45%)

Pedant information for DKFZphtes3_17f10, frame 3

Report for DKFZphtes3 17f10.3

```

SEQ      MDRSQQTSRTGVTWNNMI PPVEKVQDEQQTYFSSEITVVISRPDSSSTKSKEDALKHKSS
SEG
PRD      cccccccccccccccccccceehhhhhhhcccccceeeccccccccchhhhhhhhhccc

SEQ      GKFASHEHPFQPATNSNEEI GQKNI SRTSFTQETKKGPPVLELEDLREEVTVPVVQEGS
SEG
PRD      cceecccccccccccccccccccccccccceeeccccccchhhhhhhhhhheeecccccc

SEQ      AVKKVASAEIEPPSTEKFPKAIQPLVEEATAKAEPPRAETHVQPSQSTTEPDAEAET
SEG
PRD      chhhhhhhccccccccccccccccchhhhhhhhhccccccccceeeccccccccchhhhhh

SEQ      AXAENSVKVQPPPAEAPLVEFPAEIQPPSAEESPVSELLAEILPPSAEESPEEPFAEI
SEG
PRD      hhhhhcccccccccccccceeeccccccccccccchhhhhhhcccccccccccccccccc

SEQ      LPPPAECPSPVELLGEIRSPSAQKPIEVQPLPAEGALAEAPKVEPTVEETLAEVQPL
SEG
PRD      xxxxxxxccccccccccccccccccccccccccccchhhhhhhccccccccccccchhhhhhh

SEQ      LPEEAPREEARELQSTAMETPAEAEAPFQSPLEPKETAEAEASAIQLLAATEPPADET
SEG
PRD      xxxxxxxxxxccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccc

```

```
SEQ PAEARSPLSEETSAEEAAHVQSPLAEETTAEASAEIQLLAAIEAPADETPAEAQSPLS
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccc

SEQ EETSAEAPAEVQSPSAGVSIIEAPLELQPSGEETTAEASAAIQLLAAETASAEAP
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD chhhhhccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ AEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAPAEVQPPPAEAP
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AEVQPPPAEAPAEVQPPPAEAPAEVQSPAEVSLPAETPIETIELAAVHSPADDVPAEASVD
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc

SEQ KHSPPADLLLTIEFPFGEASAEVSPFSEQTPEDALVENVSTFQSPQVAGIPAVKLGS
SEG cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc
PRD cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc

SEQ VVLEGEAEFEVSVKSLVDKQNTDQOAPLTLESVFIHELCKKQRPPEL
SEG eeehhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhcccccc
PRD eeehhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhcccccc
```

(No Prosite data available for DKFZphtes3_17f10.3)

(No Pfam data available for DKF2phtes3_17f10.3)

DKF2phtes3_17117

group: metabolism

DKF2phtes3_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```
1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51  GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGGCAGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTGCTGT CTGTCTCTCT CTCCACACG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACACAG ACCGGTTTAT CCTCTCCAGG GGACATGCTG
251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGAGAA GACACCCCTAC
351 CCCGCGATTG CCGTTTGTGT ACGTGGCAAC AGGGTCCCTA GGTGAGGGAT
401 TAGGTACTGC ATGTGGAATG GCTTATACGT GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCTCT TATGGGAGAT GGCGAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGCTT TTGCCTCCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
601 GAGCATGGCG CAGACATCTA CCAGAAATGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGGTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATAC CAAAATCGCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCTTA
951 CAAAGTTGGT GACAAGATAG CTAATCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAGCTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAACCTCA CCTTTTCTGA GATATTCAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAAGTGT GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTGATGCA GCTCCGAATG GGAGCCATTG CTCAGGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGGCCTGGAG GATCTAGCCA TGTCCGAAG CATTCCTAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTGGAACC AGCCAAACAG
1451 AAAGTGCAGT TATTTATACC CCACAAGAAA ATTTGAGAT TTGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTCACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTTCTGT CCGTGTATC GACCCATTTA CCATTAAACC CTTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GGCGCCGAG TTATCACAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTTGTTATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAATACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1901 TTATTTCTAA AAGTCAAGT CTATTGGCTT TGGCCCAAAA GCACTGGTAT
1951 CTTTGTATTA AATTCATGTT TATTGTCACA AAACCATTAT TTATACCTAT
2001 ACAGTTGTAC TGTTCCTTTT AAAGCAAGC CATTAAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAG TTTACCTTTT TGTTAATCTA TGTATAAATG
2101 TTAATCTGAG TTATTAATGT GGAATTTAAA ATGTAAAGCA ATAGAATAGG
2151 AAATAAAACA ACTACCTAAT ACAAAATATT CTGATAAGAC TACAAATATC
2201 TGACTGAGCT GGGGATTAAG GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCTCTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTGT AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTACAG ATAAATGTTT ACTGCAGTTG CTTGGAAAT TCCTCCAAAG
2401 TTTGCCCTCA TCTCTCTCTT ACAGTTTGGG GGTGATGGTG CAGCAGTGGG
2451 ACATCTCTTG ATGCACCACA CTACTTGTGT TCTGTGAAGT GATGAAAGTA
```

2501 TAACTGGTTC TAGTTTGCAC ACTACACACA TAGTTTGTG AAGCTTCAGA
2551 AATGTTTTTT CTTTTCCTTG TGGCCAAACC AGTTTGTAA TCTGATTATA
2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAGCAT TTAATAATCA
2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

96214928:
Amplification of the transketolase gene in desensitization-resistant
mutant
Y1 mouse adrenocortical tumor cells.

99123875:
Properties and functions of the thiamin diphosphate dependent enzyme
transketolase.

Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626
Category: strong similarity to known protein
Classification: Metabolism
Prosites motifs: ATP_GTP_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS
51 VLFHTMKYK QTDPEHPDND RFILSRGHAA PILYAANVEV GDISEDDLN
101 LRKLHSDLER HPTPRLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV
151 FCLMGDCESS EGSVWEAFAP ASHYNLNLV AVEDVNRLLGQ SGPALEHGA
201 DIYQNCCEAF GWNTYLVDPGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR
251 GIPNIEDAEN WHGKVPVKER ADAIVKLIES QIQTNENLIP KSPVEDSPQI
301 SITDIKMTSP PAYKVGDKIA TOKTYGLALA KLGRANERVI VLSGDTMNST
351 FSEIFRKEHP ERFIECIIAE QNMVSVLALGC ATRGRITIAFA GAFAAFETRA
401 FQQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRISPNCTVF
451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR
501 HGVNDKVTVI GAGVTLHEAL EAADHLSQQG ISVRVIDPFT IKPLDAATII
551 SSAKATGGRV ITVEDHYREG GIGEAIVCAAV SREPDILVHQ LAVSGVPQRG
601 KTSLELDMPG ISTRHIIAAV TLTLMK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_17117, frame 1

SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68)., N = 1,
Score = 2222, P = 2.5e-230

SWISSPROT:TKT_RAT TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score =
2202, P = 3.3e-228

TREMBL:RN09256_1 product: "transketolase"; Rattus norvegicus
Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202,
P = 3.3e-228

SWISSPROT:TKT_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score =
2200, P = 5.3e-228

>SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).
Length = 623

HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230
Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFFHTM+YK DP +
 6 KPDQKQLQALKDNTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYAAMVEVGDISESDDLNLRLKLSDLERHPTPRLPFVDVATGSL 126
 P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLGDHPVPKQAFDQVATGSL 125

Query: 127 GQGLGTACGMAYTGKYLKASYRVFLMGDGESEGSVWEAFASFASHYNLDNLVAVFDVN 186
 GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA Y LDNLVA+FD+N

Sbjct: 126 GQGLGAACGMAYTGKYFDKASYRVYCHLGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGSPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246
 RLQGS PAPL+H DIYQ CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQGS DPAPLQH QVDIYQKRCEAFGWHTIIVDGH SVEELCKAFGQA---KHQPTAI AKT 242

Query: 247 FKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTENLIPKSPVEDSPQISITDIK 306
 FKGRGI IED E WHGKP+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEIYSQVQSKKILATPPQEDAPSVDIANIR 302

Query: 307 MTPSPAYKVGDKIATQRTYGLALAKLGRANERVIVSGDTMNSTFSEIFRKEHPERFIEC 366
 EDG SQMALEDLAFMR+P TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 303 MTPSPSYKVGDKIATRKAYGLALAKLGHASDRIIALDGDTKNSTFSEIFRKEHPRDFIEC 362

Query: 367 IIAEQNMVSVAGCATRGRTIAFAGFAFAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426
 IAEQNMVS+A+GCATR RT+ F FAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSIAGCATRDRTVPFCSTFAFFTRAFDQIRMAAISNINLIGSHCGVSGIS 422

Query: 427 EDGVSQMALEDLAFMRIPNCTVFYPSDAISTEHAIIYLAANTKGCMCFIRTSQPETAIVIY 486
 EDG SQMALEDLAFMR+P TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAFMRVPMSTVFYPSDGVATEKAVELAANTKGCFIRTSRPEAIYIS 482

Query: 487 PQENFEIQQAQVVRHGVNDKVTIVIGAGVTLHEALEAADHLSQGGISVRVIDPFTIRPLDA 546
 E+F++GQAQVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIRPLD

Sbjct: 483 NNEDFQVQAQVVLKSKDDQVTIVIGAGVTLHEALAAEESLKKDKISIRVLDPFTIRPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYREGGIGEAACAASREPDIIVHQLAVSGVPQRGKTSSELL 606
 I+ SA+AT GR++TVEDHY EGGIGEA AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYREGGIGEAASAAGVGEFVTVTRLAVSQVPRSGKPAELL 602

Query: 607 DMFGISTRHIIAAV 620
 MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKF2phtes3_17117, frame 1

Report for DKF2phtes3_17117.1

[LENGTH] 626
 [MW] 67877.52
 [PI] 5.90
 [HOMOL] SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0
 [FUNCAT] m outer membrane and cell wall [M. jannaschii, MJ0681] 3e-48
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32
 [FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17
 [FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09
 [FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
 [BLOCKS] BL00801F
 [BLOCKS] BL00801E
 [BLOCKS] BL00801D Transketolase proteins
 [BLOCKS] BL00801C Transketolase proteins
 [BLOCKS] BL00801B Transketolase proteins
 [BLOCKS] BL00801A Transketolase proteins
 [SCOP] dltkka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domain 1e-21
 [EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11
 [EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10
 [EC] 2.2.1.1 Transketolase 0.0
 [EC] 2.2.1.3 Formaldehyde transketolase 1e-20
 [PIRKEW] transferase 0.0
 [PIRKEW] flavoprotein 2e-07
 [PIRKEW] Calvin cycle 1e-40
 [PIRKEW] heterotetramer 2e-07

```

[PIRKW]    pentose phosphate pathway 0.0
[PIRKW]    magnesium 1e-40
[PIRKW]    thiamine pyrophosphate 0.0
[PIRKW]    oxidoreductase 7e-12
[PIRKW]    fatty acid biosynthesis 4e-10
[PIRKW]    mitochondrion 2e-07
[PIRKW]    peroxisome 1e-20
[PIRKW]    homodimer 1e-40
[SUPFAM]   pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
[SUPFAM]   pyruvate dehydrogenase (lipoamide) beta chain 7e-12
[SUPFAM]   ferredoxin 2[4Fe-4S]-related protein 8e-47
[SUPFAM]   thiamine pyrophosphate-binding domain homology 0.0
[SUPFAM]   pyruvate dehydrogenase (lipoamide) 6e-08
[SUPFAM]   ferredoxin 2[4Fe-4S] homology 8e-47
[SUPFAM]   hypothetical protein C2814 2e-21
[SUPFAM]   transketolase 0.0
[PROSITE]  ATP_GTP_A 1
[PFAM]     Transketolase
[KW]       Alpha_Beta
[KW]       3D
[KW]       LOW_COMPLEXITY 3.04 %

```

```

SEQ  MMANDAKPDVKTQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYK
SEG  .....
1ngsB ..... HHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHH-HHCCCT

SEQ  QTDPEHPDNDRFILSRGHAAPILYAAWVEVDISEDLLNLRKLSDLERHPTPLRPFVD
SEG  .....
1ngsB TTTTTTTTCEEEETTGGGHHHHHHHHHHHCTTCHHHHHHTTTTTTTTTTTTTTTTTTTC

SEQ  VATGSLGQLGTACGMAYTGKYLKASYRVFCLMGDGESSEGSVWEAFASFASHYNLDNLV
SEG  .....
1ngsB CCCCTTHHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHCHHHHHHHHHHHHHHCTTTEE

SEQ  AVFDVNLGQSGPAPLEHGADIYQNCCEAFGWNTYLVGDHVEALCQAFWQASQVKNKPT
SEG  .....
1ngsB EEEEECEETTEEAGGCCCHHHHH-HHHCEEEETTTTHHHHHHHHHHHHHHTTTTCE

SEQ  AIVATFKGRGIPNIEDAENWHGKVPVKERADAIVKLIESQIQTENLIPKSPVEDSPOI
SEG  .....
1ngsB EEEEECTTTTTCCHHHHHHHHHHNTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ  SITDIKMTSPPAYKVGDKIATQKTYGI-ALAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
SEG  .....
1ngsB HHHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHHHTTTTTTEEEETTTTHHHHCCCTCECCCG

SEQ  ERFIECIIAEQNMVSVLGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISOANINLIGSH
SEG  .....
1ngsB GCEETTTTHHHHHHHHHHHHHHTTTTEEEEGGGGGGHHHHHHHHHHHCTTTEEEEC

SEQ  CGVSTGEDGVSQMALEDLAFRSIPNCTVFYPSDAISTEHAIFYLAANTKGMCFIRTSQPE
SEG  .....
1ngsB CCGGGTTTTTTTTCCHHHHHHHHCTTTTEEECCCHHHHHHHHHHHHTTTTCEEECCCCCB

SEQ  TAVIYTPQENFEIGQAKVVRHGVNDKVTIVIGAVTLHEALEADHLSQOGISVRVIDPFT
SEG  .....
1ngsB CCTTTTCHHHHHCC-CEEEETTTTTEEEECCHHHHHHHHHHHHHHHHHHCCCEEE...

SEQ  IKPLDAATIISSAKATGGRVITVEDHYREGGIGEAVCAAVSREPDIIVHQLAVSGVPORG
SEG  .....
1ngsB .....

SEQ  KTSELLDMFGISTRHIIAAVTLTLMK
SEG  .....
1ngsB .....

```

Prosite for DKFZphtes3_17117.1

PS00017 595->603 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_17117.1

HMM_NAME Transketolase

HMM *vNtIRiLaMDAVEKANSGHGPGaPMGMAPMAHVLWqrMMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL++++M++++DP P+ ANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYKQTDPEHPD	68
HMM		RDRFVLSNGHacMLLYsMWHLYGYDmPMWDLkQFRQWHSrTPGHFEIghT +DRF+LS GHa+++LY+ W + G +++++DL+++R++HS++ +HP ++	
Query	69	NDRFILSRGHAAPILYAAWVEVGD-ISESDLNLRKLHSDLERHPTPLRP	117
HMM		PGVEVTTGPIGGQIaNaVWMAIAERNLAATYNRPFGDIFDHYTYCFMGDG ++ +V+TG+LGQG++ +++++Y++++ D+++++C+MGDG	
Query	118	FV-DVATGSLGQGLG-----TACGMAYTGKYLDKASYRVFCLMGDG	157
HMM		CLMEGISWEACSLAGHMqLGNWiaFYDDNcISIDGdTdIWfQEdtYakRF + +EG++WEA ++A++H++L+N++A +D NR++++G++++ + D+Y+ +	
Query	158	ESSEGSVWEAFaFASHYNLDNLVAVFDVNRLGQSGPAPLEHGADIYQNC	207
HMM		EAYGWHVIEVENdGHdVeeIcaAIEeAKaekDRPTLIICRTVIGYGSNPK EA+GW++ +V DGHdVE++C A+ +A +K++PT+I ++T++G+G+PN	
Query	208	EAFGWNTYLV--DGHdVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI	255
HMM		QGTHdWHGAPLGeD* ++ + WHG+P +++	
Query	256	EDAENWHGKVPKE 269	
HMM		*PqWePnddKIATRKASQqaLeaiGPaLPefWGSADLTSPNLtWKgav P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S+++++ ++	
Query	311	PAYKV-GDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKE	358
HMM		WFMPPSISTDCynGNWwGRIYHYGIREHGMaAIMNGIALHGgNFRPYGGT + + R+I++ I+E+++M++++ G+A++G+ +++++ G	
Query	359	H-----PERFIECIIAEQNMVSVLGCATRGR-TIAFAGA	392
HMM		FMMfyDYARPAIRMAALMelPVIWVTHDSIGLGEDGPTHQPVENLAHER F+++F+++A++++RM A++ +++++H++++ GEDG +++++E+LA+ER	
Query	393	FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSQMALEDLAMFR	442
HMM		aIPNMwVWRPCDgNETayAWylavERhTPtLiILSRQNLQLErNPrqf +IPN v+++P+D+ T+ A YLA++++ +++++S ++ +++++ P +	
Query	443	SIPNCTVFYPSDAISTEHAITYLAANTKGM-CFIRTSQPETAIVIY-PQEN	490
HMM		ekvaRGGYVLkDmdnePDVILIATGSEMELavaAAKlLadEGIKaRVVSM +++++++V + + + V++I++G++++A++AA+ L+ +GI +RV+++	
Query	491	FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQQGISVRVIDP	538
HMM		PCTeWFD.....kQDeEYReSVLPdhVPqRvAVeMgvtWCWYKYVGqq +++++D +++++R ++DH++ +++++V ++ +++ +	
Query	539	FTIKPLDAATIISAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDL	587
HMM		GaIfGMN:FGESSGRAPpevLYkMFGFTPENI* + +++ +++ ++ +L+ MFG+ +I	
Query	588	VHQLAVSGVPQR---GKTSELLDMFGISTRHI 616	

DKFZphtes3_l7n12

group: transcription factors

DKFZphtes3_l7n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucin-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```
1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGTCTCAT GTAACAATAG TAGATTGTTT TTTTTCCTAA TATTTCTAGC
101 CAGCCCTCTAA GTCAGGTGAT GGAACAAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GGTGGAGGAA GGAAGAAGA AAAATGGGAA GAATGTCTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCTGTGTC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTAACTCTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
301 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
351 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
401 ACAGCGTTCT GTCATCTCAG CAAGAATGGG AATCAGAGAT GCTGACTGGG
451 TGTTCCTTAT ATTCTCTCCG AAATACCTCT ACCTCACCAC ATAAGCTCTG
501 CGAAGGGAGT CGGGACCGGT AGATAATGAC CAGTGTACTT TTTGGAAACC
551 CAGAGCGCGG CAAGAAGGAT CTGCGCATGC TGGTCGACAC ACTGAAACAG
601 AAGAAGCTTG AGGAATGAT TCGGACTGAA CAAGAGGATT CCTCCTGCAT
651 GGAAGAACTA CTTTCAAAAG ATTGGAAGGA AAAAATGGA AGACTAAATA
701 CCAGTAACTA CTTTCAAAAG ATTGGAAGGA AAAAATGGA AGACTAAATA
751 AAGAAGCGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTCACGGA
801 GCAGCTACTG CGAGCGCATG ATGAACAGAA AAAACTGGCA CGCTCACAAA
851 TTCAGAAACA ACGGACGCAA ATGGACCTTG CTGCGCAACA GCAAGAACAG
901 ATTCCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCCTCCG CTCATGATCC
1001 CAATTTTTC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGATATCC TCTTCCCTCC TGGAAATACA TACAAACCAG GTGATAACTA
1101 CCCCCTACAG TTCATTCCAT CAACAATGGC AGTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTCTC CTGGAGCAAA GATGCCATCA ACTCCACAGC CACCAAAAC
1251 AGCAGGAGCG GTCTCACCTA CTGGGATAAA AAATGAAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCACAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGACAGC AGAGCCTGTA AAGTCCCAAA CGTCTCCAC
1401 CCAGAACCTC TTCCAGCCCA GCAAAACCAG CCCTGTCAAT CTGCCAAAAC
1451 AAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAGTCA ACACAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCACTCCC CTGCCCCTTT TGGGGATCAG GATACAGTGA
1601 TGAAGCCAT TCAGGAGGCG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTGT
1751 AGAATTTGGG GCCCGATTG ACGGGAAAGT CAAATGAAGA TGGAAAAGTG
1801 GGCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGAGC CTAACTACA GCAGTATTAT TGTGGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCAGCAG TCTACAGGGA CGCCCGCGGC
1951 CGTGCAGCA GCGAGCCACA CATTAGCGGA CCAATGAATG CATTCTATGT
2001 TTGGGCAAGG GATGAGGAGA GAAAAATCCT TCAGGCCCTTC CCCGACATGC
2051 ATAACCTCAA CATTAGCAAA ATCTTAGGAT CTGCTGGGAA ATCAATGTCC
2101 AACCCAGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAACCCCGCA CCGAAACGCA
2201 CCTGCATTGT TGATGGCAAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGCAG TTCTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACA CAGGAACAGG TGTGTGTAT CCTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGTCTT
```

```

2401 AGCACCTCGG CCAGCCCGGA GCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAATG TATGATGACT ATGAAGATGA CCCCAATCA
2551 GACTATAGCA GTGAAATGA AGCCCGGAG GCTGTCAGTG CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCA
2651 ACAAGAGATT ATTAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

BLAST Results

No BLAST result

Medline entries

95311974:

A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:

The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804
Category: strong similarity to known protein

```

1 MGRMSSKQAT SPFACAAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESNNKLC SLYSFRNTST
101 SPHKPDEGSR DREIMTSVTF GTPERRKGS L ADVVDTLKQ KLEEMTRTEQ
151 EDSSCMEKLL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMITQ
201 LLSLREQLLA AHDEQKKLAA SQIEKQRQDM DLARQQEQEI ARQQQQLLQQ
251 QHKINLLQQQ IQVQGHMPPPL MIPFPHDQR TLAAAAAQQ GLFPPPGITY
301 KPGDNPVPQF IPSTMAAAAA SGLSPLQLQQ LYAAQLASMQ VSPGAKMPST
351 PQPNTAGTV SPTGIKNEKR GTSPTVQVKD EAAQPLNLS SRPKTAEPVK
401 SPTSPTQNL FASKTSPVNL PNKSSIPSP I GGLGRGSSL GKWKSOHQEE
451 TYELDILSSL NSPALFGQDD TVMKAIQEAR KMREIQREQ QQQPHGVGD
501 KLSSINNMGL NSCRNEKERT RFENLGFQLT GKSNEGKLG PGVIDLTRPE
551 DAGSKAMNG SAAKLQYYC WFTGGATVAE ARVYRDARGR ASSEPPIKRP
601 MNAFMVWAKD ERRKILQAFP DHMNSISKI LGSRWKSMN QEKQPYVEEQ
651 ARLSKIHLK YPNYKIKPRP KATCIVDGKK LRIGETKQLM RSRROEMROF
701 FTVGQQPQIP ITTGTGVVYP GAITMATTTT SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMERY DDYEDDFKSD YSSENEAPEA
801 VSAH

```

BLASTP hits

Entry MMSOXL22.1 from database TREMBL:
product: "SOX-L2"; Mouse mRNA for SOX-L2, complete cds.
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:
SOX-L2 - rainbow trout
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:
SOX6 protein - mouse
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330.1 from database TREMBL:
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604.1 from database TREMBL:
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281

Alert BLASTP hits for DKFZphtes3_17n12, frame 1
No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n12, frame 1

Report for DKFZphtes3_17n12.1

[LENGTH] 804
[MW] 89332.69
[pI] 6.97
[HOMOL] TREMBL:MMSOXLZ2_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07
[FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06
[FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04
[SCOP] dlhmf_1.20.1.1.1 HMGI, fragments A and B [rat/hamster (Rattus 1e-13
[SCOP] dllefa_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mous 4e-15
[SCOP] dlhrya_1.20.1.1.4 SRY [Human (Homo sapiens) 7e-17
[PIRKB] DNA binding 4e-94
[PIRKB] T-cell receptor 4e-07
[PIRKB] leucine zipper 1e-38
[PIRKB] alternative splicing 2e-07
[PIRKB] transcription factor 4e-16
[PIRKB] transcription regulation 1e-12
[SUPFAM] HMG box homology 0.0
[SUPFAM] unassigned HMG box proteins 4e-94
[PROSITE] ATP_GTP_A 1
[PROSITE] LEUCINE_ZIPPER 1
[PROSITE] MYRISTYL 6
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 6
[PFAM] HMG (high mobility group) box
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 13.81 %
[KW] COILED_COIL 3.48 %

SEQ MGRMSSKQATSPFACAAADGEDAMTQDLTSREKEEGSDQHVASHLPHPIMHNKPHSEELP
SEG
COILS
lnhm-
SEQ TLVSTIQQADWDVLSQQRMESNNKLCSLYSFRNTSTSPHKPDGSRDREIMTSVTF
SEG
COILS
lnhm-
SEQ GTPERRKGLADVVDTLKQKKLEEMTRTEQEDSSCKEKLSSKDWKEKMERLNTSELLGEI
SEG
COILS
lnhm-
SEQ KGTPESLAEKERQLSTMITQLISLREQLAAHDEQKKLAASQIEKQRQOMDLARQQQEQEI
SEG
COILS
lnhm-
SEQ ARQQQQLLQQQHKNLLQQQIQVQGHMPLMIPFPHDQRTLAAAAAQQGFLFPPGITY
SEG
COILS
lnhm-
SEQ KPGDNYPVQFIPSTMAAAAAGLSPLQLQQLYAAQLASMQVSPGAKMPSTPQPNTAGTV
SEG
COILS
lnhm-

```

COILS .....
lnhm- .....

SEQ      SPTGIKNEKRGTSPTVQVKDEAAQPLNLSSRPKTAEPVKSPSPTONLFPASKTSPVNL
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PNKSSIPSPIGGSLGRGSSLGKWKSOHQEETVELDILSSLSNPFQDQDTVMKAIQEAR
SEG      .....
COILS    .....
lnhm-    .....

SEQ      KMREQIQREQQQQPHGVDGKLSSINNMLNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PGVIDLTRPEDAEGSKAMNGSAAKLQYYCWP TGCGATVAEARVYRDARGRASSEPHIKRP
SEG      .....
COILS    .....
lnhm-    .....CCC

SEQ      MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEKPYYEEQARLSKIHLEK
SEG      .....
COILS    .....
lnhm-    .....CCCHHHHHHHHHHHHHHTTTTCCHHHHHHHHHHHTTTTTHHHHHHHHHHHHHHHHHHH

SEQ      YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRQEMRQFETVGGQQPIPIITGTGVVYP
SEG      .....
COILS    .....
lnhm-    .....HHHTTTTTTT

SEQ      GAITMATTTPSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGS LAGNEMINGEDEMEMY
SEG      .....
COILS    .....
lnhm-    .....

SEQ      DDYEDDPKSDYSSSENEAPEAVSAN
SEG      .....
COILS    .....
lnhm-    .....

```

Prosites for DKF2phtes3_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKF2phtes3_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWQEMRekIKaENPNdMhNtEISKMiGEMWKnMsEEEkM +KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISILGSRWKSMSNQEKQ	644
HMM	PYEdMAeeEKqRYMKEMPeYK*		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKFZphtes3_17n18

group: intracellular transport and trafficking

DKFZphtes3_17n18 encodes a novel 782 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein containses TONB_DEPENDENT_REC_1 Pattern and ATP_GTP_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp

Poly A stretch at pos. 2806, no polyadenylation signal found

```
1  GTCCTTTTAA  GTCAGTAAAT  TGAAC TAAGT  CGGTTATTCG  GCAAGCAGTT
51 CCTATAAAAA  ACTACATGGC  TAAGTTCTT  AATGATTGAC  CACAAGCAGA
101 TCTTTTCACCC  TCGGATCTCT  AGCTACAAAA  GGTCCCCACA  CTGAAGAAGC
151 CACTACCTCC  ACCACCACCA  GCACCACCAC  GTCCAGTGCT  GCTGGCAACC
201 ACTGGGGCAG  CCAAGCGCTC  CACCCTCTCT  CCCACCATGG  CCGTCAGGT
251 GCGCACCAC  CAGGAGACCC  TGAACAGGTT  TCAGCAGCAG  TCCATCCACC
301 TGCTGACGGA  GCTCCTCAGA  CTGAAGATGA  AGGCCATGGT  GGAGTCTATG
351 TCGGTGGGTG  CCAACCCCTT  GGACATCACC  AGGCGCTTTG  TGGAGGCCAG
401 CCAGCTCCTC  CACCTCAATG  CCAAGGAGAT  GGCCTTCAAC  TGCCTGATCA
451 GCACAGCCGG  GAGAAGTGCC  TACAGCAGCG  GACAGTTGTG  GAAAGAGTCC
501 CTCGCAAAACA  TGTCCGCCAT  TGGGGTGAAC  TCGCCTTACC  AGCTGATCTA
551 CCACCTCTCC  ACAGCCTGTC  TGAGCTTTTC  TCTCTCTGCT  GGAAAAAGAG
601 CCAAGAAGAA  AATAGGCAAA  TCTAGAACTA  CAGAAGATGT  CAGCATGCCG
651 CCCCTGCATC  GAGGAGTGGG  AACCCCTGCC  AACAGCCTGG  AGTTCAGCGA
701 CCCCTGCCCT  GAGGCCCGGG  AGAAGCTGEA  GGAGTTGTGT  CGCCACATAG
751 AAGCTGAAAG  GGCCACATGG  AAAGGGAGGA  ATATCTGCTA  CCCCATGATC
801 TTACGAAACT  ACAAGGCAAA  GATGCCCTCT  CATCTAATGT  TGGCCCGCAA
851 AGGAGACTCT  CAGACCCCGG  GTTTACATTA  CCTCCCACT  GCAGGTGCTC
901 AGACTCTCAG  CCCCACTCT  CACCCATCTT  CTGCCAACCA  TCATTTCAGT
951 CAGCATTGTC  AAGAGGGGAA  GGCACCCAA  AAGCCCTTCA  ACTTTCATTA
1001 CACCTTCTAT  GATGGCTCCT  CCTTCGTTTA  CTATCCCTCT  GGAACGCTCG
1051 CTGTATGTCA  GATCCCAACA  TGCTGCAGAG  GGAGAACCAT  CACCTGCCTC
1101 TTTAATGACA  TACCTGATTA  CTCTTGCTGT  GCCCTATTCA  ATACTGAAGG
1151 CCAGGGCTGT  GTTCACTACA  ACCTAAAAAC  CAGTTGCCCA  TATGTCTTAA
1201 TCTTGGATGA  GGAAGTGGG  ACCACCAATG  ACCAGCAGGG  CTATGTAGTC
1251 CACAAGTGGG  GCTGGAATTC  CAGGACAGAG  ACCCTGCTTT  CCTGGAATA
1301 CAGAGTGAAT  GAGGAAATGA  AACTAAAGGT  ACTGGACAG  GACTCCATCA
1351 CAGTCACCTT  CACCTCCCTG  AATGAGACAG  TAACACTCAC  TGTGTCGGCC
1401 AACATTTGTC  CCCATGGAAT  GGCATATGAC  AACCGGCTGA  ACCGCAGAAT
1451 CAGCAACATG  GACGACAAGG  TGTATAAGAT  GAGCCGAGCC  CTGGCTGAGA
1501 TCAAGAGCGG  GTTTCAGAAG  ACAGTGACTC  AGTTCATTAA  TTCTATCTTG
1551 CTGGCCGCAG  GTCTGTTTAC  CATTGAATAT  CCCACCAAAA  AGSAGGAGGA
1601 AGAATTTGTT  CGGTTCAAGA  TGAGATCCAG  AACTCATCCC  GAGCGGCTCC
1651 CCAAGCTAAG  TTTATACTCA  GGAGAAAGTC  TTTTACGATC  TCAGTCAGGC
1701 CACCTGGAAT  CCTCAATTGC  AGAGACTTTG  AAGGATGAGC  CTGAGTCTGC
1751 TCCTGTGAGC  CCAGTTCGGA  AGACCACCAA  AATCCACACC  AAAGCCAAAG
1801 TCACATCCAG  AGGGAAGGCC  CGCGAGGGGC  GCAGCCCCAC  CAGGTGGCGG
1851 GCCTTGCCCT  CAGACTGCCC  GCTGCTGCTG  CGGAAGCTCA  TGCTCAAGGA
1901 AGACACCCGT  GCTGGTGCA  AGTGCTGGT  GAAGGCCGCC  CTGCTCTCTG
1951 ACGTGGAGCT  GGAGCGGTTT  CTGTTGGCGC  CCGGAGACCC  CAGCCAAAGT
2001 CTGGTGTGTT  GGATCATCTC  AAGCCAGAAC  TACACCAGCA  CTGGGCGAGT
2051 CCAGTGGCTG  CTGAACACTC  TCTACAACCA  CCAGCAGCGG  GGCCGTGGCT
2101 CCCCTTGCT  CAGTGCCTGC  TATGACTCCT  ACCGCTGCTG  GCAGTATGAC
2151 CTGGACAGCC  CCCTGCAGGA  GGACCTCTCC  CTGATGGTGA  AGAAGAACTC
2201 TGTGGTGCAG  GGGATGATTC  TGATGTTTGC  CGGGGGGAAG  CTCATTTTGT
2251 GGGGCCGTGT  TTTGAATGGA  TATGGCTTCA  GCAAGCAGAA  TCTGCTGAAA
2301 CAGATCTTCC  GGTCTCAACA  GGATTACAAG  ATGGGCTACT  TCCTGCCGGA
2351 TGACTACAAA  TTCAGTGTTT  CCAACTCTGT  CCTGAGCCTG  GAGGATTCTG
2401 AATCAGTCAA  GAAAGCCGAG  TCAGAAGATA  TCCAAGGAAG  CAGCTCCTCA
```

WO 01/12659

PCT/IB00/01496

```
2451 TTGGCCCTGG AAGACTATGT GGAGAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG GCCTCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGGCCGGGGT GCTGGGGCTT CTGCCCAGCC
2651 CAGCCCTGCC TCCCCGGTCT CCCACCCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGGC TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGCGCG
2851 CCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782
Category: putative protein
Prosite motifs: ATP_GTP_A (122-130)
TONB_DEPENDENT_REC_1 (1-44)

```
1 MARQVRTHQE TLNRFQOQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSGYS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSTA CLSFSLSAGK EAKKIKGKSR TTEDVSMPLP HRVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTTAG AQTLSPTSHP SSANHHFSQH COEGKAPKKA
251 FKPHYTFYDG SSFVYYPGSG VAVCQIPTCC RGRITITCLFN DIPGFSLLAL
301 FNTEGQGCYH YNLKTSCPYV LILDEEGGTT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLKVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRISNMDD KVKMSRALA EIKKRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEFVRP KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTHHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLAK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPQVQLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCIQCRDY SYRLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSRQNLKQI FRSQDDYKMG
701 YFLPDDYKFS VPHSVLSLED SESVKRAESE DIQSSSSSLA LEDYVEKELS
751 LEAENTREPE VELHPLSRDS KITSWKKQAS KK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17n18, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3_17n18, frame 3

Report for DKFZphtes3_17n18.3

```
{LENGTH} 782
{MW} 88030.16
{pI} 9.22
{BLOCKS} BL00286 Squash family of serine protease inhibitors proteins
{PROSITE} ATP_GTP_A 1
{PROSITE} MYRISTYL 4
{PROSITE} CAMP_PHOSPHO_SITE 3
{PROSITE} CK2_PHOSPHO_SITE 14
{PROSITE} PROKAR_LIPOPROTEIN 1
{PROSITE} TONB_DEPENDENT_REC_1 1
{PROSITE} PKC_PHOSPHO_SITE 10
{PROSITE} ASN_GLYCOSYLATION 4
{KW} Alpha_Beta
```

SEQ	MARQVTRHQTETLNRFFQOQSILHLTELRLRLMKMAMVESMVSQANPLOTIRRFVEASQLLHL
PRD	ccccchhh
SEQ	NAKEMAFLNCISITAGRSVGSYLKWKSSLANMSAIVGNVSPQLIYHSSTACLFSLSLACK
PRD	hhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccc
SEQ	EAKKIKGSRRTTEDVSMPLLRHGVTFPANSLFSDPCPEARKLQELCRHIEAERATWKG
PRD	hh
SEQ	RNISYPMILRNYKAKMPSHMLARKGDSQTPGLHYFPPTAGATLSPTSHSSANHHFSQH
PRD	ccccchhhhhhhhhhhhhcc
SEQ	CQEGKAPKKAFFHYTIFYDGSSFVYYPGSGNAVQCITPCRCGRITICTLFDIPGFSLLAL
PRD	ccccchhh
SEQ	FNTGEGQGVHYNLKTCSPYVILDEEGGTTNDQQQGVVHKWSWTSRTETLLSLEYKVNEE
PRD	ccchhhhhhhhhhhhhhh
SEQ	MKLVLHGQDSITVTFTSLNELVTLTYSANNCPHGMAYDKRLNRRISNMDDCKVYKMSRALA
PRD	hh
SEQ	EIKKRFQKTVTQFINSILLAAGLFTIEYPTKKEEEFVRFKMRSRTHPERLPKLSLYSGE
PRD	hh
SEQ	SLLRSGQGHLESSIAETLKDEPESAPVPRKTTKIHTKAKVTSRGKAREGRSPTRMAAL
PRD	EEEECCCCCHHH
SEQ	PSDCPLVLRLMLKEDTRAGCKCLVRKPLVSVDELFRHLFLHAPDPQSQVLFVFGISSQNT
PRD	ccccchhh
SEQ	STGOLQWLNTLYNHQGRGSPCIQCRYSYRLLYQDLDSFLQEDPPLVMKKNVSVQGM
PRD	chhh
SEQ	ILMFAGGKILFGGRVLNGLSKQNLKQIFRSQQOYKMGYFLPDDYFVSFVNSVLSLED
PRD	HEVEKCCCCCCCCCCCCCCCCCHHH
SEQ	SESVKKAESDEDQGSSELALEDYVEKELSLEAKTREPVELHLPLSRDSKITSWKKQAS
PRD	chhh
SEQ	KK
PRD	cc

Prosite for DKFZphtes3_17n18.3

PS000001	91->95	ASN_GLYCOSYLATION	PDOC000001
PS000001	182->186	ASN_GLYCOSYLATION	PDOC000001
PS000001	379->383	ASN_GLYCOSYLATION	PDOC000001
PS000001	598->602	ASN_GLYCOSYLATION	PDOC000001
PS000004	403->407	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	511->515	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	652->656	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	48->51	PKC_PHOSPHO_SITE	PDOC000005
PS000005	177->180	PKC_PHOSPHO_SITE	PDOC000005
PS000005	344->347	PKC_PHOSPHO_SITE	PDOC000005
PS000005	450->453	PKC_PHOSPHO_SITE	PDOC000005
PS000005	497->500	PKC_PHOSPHO_SITE	PDOC000005
PS000005	513->516	PKC_PHOSPHO_SITE	PDOC000005
PS000005	523->526	PKC_PHOSPHO_SITE	PDOC000005
PS000005	631->634	PKC_PHOSPHO_SITE	PDOC000005
PS000005	723->726	PKC_PHOSPHO_SITE	PDOC000005
PS000005	774->777	PKC_PHOSPHO_SITE	PDOC000005
PS000006	7->11	CK2_PHOSPHO_SITE	PDOC000006
PS000006	131->135	CK2_PHOSPHO_SITE	PDOC000006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC000006
PS000006	329->333	CK2_PHOSPHO_SITE	PDOC000006
PS000006	345->349	CK2_PHOSPHO_SITE	PDOC000006
PS000006	377->381	CK2_PHOSPHO_SITE	PDOC000006
PS000006	406->410	CK2_PHOSPHO_SITE	PDOC000006
PS000006	450->454	CK2_PHOSPHO_SITE	PDOC000006
PS000006	466->470	CK2_PHOSPHO_SITE	PDOC000006
PS000006	493->497	CK2_PHOSPHO_SITE	PDOC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC000006
PS000006	571->575	CK2_PHOSPHO_SITE	PDOC000006
PS000006	693->697	CK2_PHOSPHO_SITE	PDOC000006
PS000006	717->721	CK2_PHOSPHO_SITE	PDOC000006
PS000008	145->151	MYRISTYL	PDOC000008
PS000008	327->333	MYRISTYL	PDOC000008
PS000008	592->598	MYRISTYL	PDOC000008
PS000008	734->740	MYRISTYL	PDOC000008

WO 01/12659

PCT/IB00/01496

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKF2phtes3_17n18.3)

DKFZphtes3_18f3

group: testes derived

DKFZphtes3_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```

1 GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51 CCCGACGCGC TGGCGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAAAG TGGCCGCGAG CTCGCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTAC
251 CCTGGGGACC TCGCTGCTGG TGTGGCCGCT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTTCTGCAAC
351 TCCCGGGAGC TCGGGAGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGGCGGA GGAACGCCCT CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCTCAT
551 CCCAGGGCGG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCTGCAC CGGGGCTCTG
651 GACGAATCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CACGACCTCA AGATCTCTGC TGACACGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAATCATCC
801 TCATGGGATG CTCAGAAATT TGTAGCTCCC TTAGGAAAAC ACCAAGCTGG
851 GTTAGGAGCC GAAGGCAAAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCG
901 GTCCCAAAAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCAAAAACCT
1001 CCTTTTCTTT TATCAAAAAC TTCTGTCTTA AACACAGCTG GGCAGGCACCT
1051 CCTGTTTAA AGTTATTTTC GGGTCCCTGA CCCTGGCCCTG GTGGCTTTGG
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTTCCCAA GTCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TGGGTGAGGA AGGCATTTCG CTCTATTCCA
1201 GAATTTCTGA TACAAGAAGC TCCAGATACC AGAGCAAAATC AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCAGAG GAGGACCCCT GTGCTGATAT
1301 CTCCTCCTCT TCCCTTTCCC CTCAGCTTAC TTACTCCAGC ATGCGGCCCTG
1351 GGTATGAAGT AGGCCTTTCC TGAGTGCTCT CCAATCCAGT CCTCAAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCTAT TAAGTCCCTG TCCCTCACAT
1451 CAAGCTGAGG GCCAGATAGA TGGAGGCACT GCCAACTTCA TTTCCCGACA
1501 TCATTGCTCT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCTGGGCG
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCTGTTT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CTGGAATGCG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAAAT
1751 CTTGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTCTCTT GGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA ATCACTACA ACAAAATGCTA ACTTCTAATC CCCTTGATGA
1901 CTTTACAGA AGTCTCAGG CTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTACAGA TGACCAAGGA CAGACTGTGT CCCAGAGCC AAATGAGAG
2001 AGAGAGAGAG AGCAGCGGTA CGTGACCCCT GGGGCAAGTG CTCACCGTAT
2051 GAATAAGGGA TGAACACTA AAAGCCCAT AGGGGGCAGT GTTTCGCCCG
2101 TGTTGTAGAA ACTGGTACAG AAAGGATCCT ATATGAAGTT CTTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAA TGGCCAGTCC ATGTATTTTT
2201 TATTTATTTT AAGTTGTAA TTTAATTTT AATTATTGTT TAGTGTGTGC
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTCTCTA GTGTTTTAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAAG
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTGCGCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAGATATA AAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAAGTTGCA AGCCTGATGT CCTATCAAGT TATGCTTCT GGGTGACAGA
2701 CAAAATAGCT TGTCTTATGG TGGTATGTTG TTGCATTTTC ACTTTGGGGT

```

```

2751 CTCTAAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTG
2801 TTAACAACCTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCTCTG TTAATAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCTT TCCAGTGAT GGGGAGATG GTCTGAGGAA
2951 GTATAATTTC AGTACTGGGG TCGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTC TATAAATTGC AATTGGTCTG TATGCTGGTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTG ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTGTAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCACTGCAA CCTCCGACTC
3201 CTGGGTTCAG GCAATTCTTC TGCTCAGCC CCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTGGGCCA GGCTGGTCTT GAATCCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGTGGG GATTACAGGC ATGAGCCACC
3401 TCGCTGGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATTGTTTTGT TTCACAATCA TTTTAAATCA TTTTAGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCCTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCATGCTT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTC TCTCATATT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATT CAGTGGCCCA CTTTATTTCT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAATCTA AGAATAACAT TTGAATAATT
3801 ATAATTAAT GTTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTCAG TTGATTGAT TCTAGAAACA AATATTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTGAAJ TGGTATGTTG TAAAAGAGAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTCTGGGG CTTCGAATCA
4151 TTTGAATTGT TCTGTTTAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATCCA ATAGAACTGC TCCATTAATA AAATAATCCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCCTTGCC CTCGGGCACC TGTCATTTC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAGAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATCCA ATATTTTAGT GTGAATTGAG ACACGTAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTCAAT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

BLAST Results

Entry HSG27587 from database EMBL:
human STS SHGC-32548.
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:
human STS EST303564.
Score = 1417, P = 8.7e-58, identities = 285/287

Medline entries

No Medline entry

Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194
Category: questionable ORF
Classification: no clue

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_18f3, frame 2

PIR:CGBOIS collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PAL_1 gene: "COL1A1": Human proalpha 1 (I) chain of type I procollagen mRNA (partial)., N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)
Length = 779

HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPAAQALPRSQRGR 62
G + G PG + AR PG GPP PA P GA AP G A A P SQ
Sbjct: 230 GDLGAPGPSGARGERGFPGERGVEGPPGAPRGANGAPGNDGAKGDAGAPGAPGSQAGP 289
Query: 63 QLAERNRGRPRRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
L G P RGA PG GD +GA G + G VR L + PG A
Sbjct: 290 GL---QGMPGE-RGAAGLPGPKGDRGDAGPKGADGAPGKDG---VRGLTGPVGGPPAG 341
Query: 123 GAGDRGHL-P-GP-----DARDPELPRVFLPLAGLRGPPAA 156
GD+G P GP D +P P P AG GPP A
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG----GPAPGAPAAQALPRSQRG 61
G G PGAA R P AGPP P P G ++G GPA G P + P G
Sbjct: 434 GATGFPGAA-GRVGGPPGSGNAGPPGPPGAGKEGSKGRGTGPA-GRPGEVGGPPGPPG 491
Query: 62 QLAERNRGRPRRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 121
A G P G PG PG RG G +RG R L PG +
Sbjct: 492 P--AGEKAGPAD-GPAGAPGTGPGQIAGQGVVGLPGQRGE---RGFPGL---PGPS 541
Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160
G +G R P P + GL GPP + RE
Sbjct: 542 GEPGKQGPSGASGERGPPG---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPAAQALPRSQRG-R 62
G G PG AR +A PG A G P A PPG + GP PG P A +G R
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATGFPGAAGRVGPPGPS-GNAGP-PGPPGAGKEGSKGR 472
Query: 63 QLAERNRGRPRRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120
GRP G + PG PG GA G G + ++ LFG
Sbjct: 473 GETGPAGRP---GEVGGPPGPAGEKAGADGAPAGAGTGPQGIAGQGVVGLPGQ 528
Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154
G+RG LPGP + P +G RGPP
Sbjct: 529 R---GERGFPLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAAWARRAAALPGT---AAGPPRPAAPPGAAPARG--GPA--PGAPAAQALPRSQR 60
G G PG + PG A+GP P PPG G G A PG P + P +
Sbjct: 29 GPPGAPGPGQPGPPGEPGEPGASGPMGPRGPPGPKNGDDGEAGKPRGPRGERGPPGPQ 88
Query: 61 G-RQLAERNRGR--RRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115
G R L G P + HRG G GD +G G G + R L
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDAGKADGAPGPKGEPSPGENGAPGQMGPRGLPGFP 148
Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
GAA G AG+RG +PGP P AG +GPP A
Sbjct: 149 GPKGAAGEPKKAGERG-VGPPGAVG--PAGKDGEAGAGGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAP--PGAAPARGGPAP-GAPAAQALPR 57
G AG PGA A PG A AGPP PA P PG G P P GA A P
Sbjct: 374 GFAGPPGADQPCAKGEPGDAGAKDAGPPGAPGAPGPPGPIGNVGAAPGPKGARGSAGPP 433
Query: 58 SQRGRQLAERNRGRPRRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G A P G PG PG +G G GR V
Sbjct: 434 GATGFPGAAGRVGPPGPSNAGPPGPPGAGKEGSKGRGTGPAGRPGEVGP----- 486
Query: 118 PGAAGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152
PG AG++G PG D A P P +AG RG
Sbjct: 487 PGPPGAGEKG-APGADGFAGAPGTGPP-GGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAARAAALPGTAGPPRPAAPPG--AAPAR-GGPAP-GAPALPRSRGR 62
 GE G A + L P G A G P P A P G P G P P G A + +RG
 Sbjct: 194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEGVPGDLGAPGPSGARGERGFPGERGV 252

Query: 63 QLAERNRPRRRHRLAQPGHFGDLAAGVGRGAGGGHRRRHHVRSLLADLLQLPGAEE 122
 + PR GA G GD A G + G +G R A L PG
 Sbjct: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPAGSGQAGPLQGMPE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157
 GDRG GP D P V L G GPP A
 Sbjct: 308 --GDRGDA-GPKGADGAPGKGV-RGLTGPIGPPGPA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03
 Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAARAAALPGTAGPPRPAAPPGAAPARG-GPAPGAPALPRSRGR 61
 NG+ GEAG PG R P A G P A PG RG GA A P +G
 Sbjct: 67 NGDDGEAGKPRRP-GERGPPPGQARGLPAGTAGLPGMKHGRGFSGLDGAAGDAGPAGPKG 125

Query: 62 RQLAE-RNGRPRRRHRLAQPGHFGDLAAGVGRGAGGGHRRRHHVRSLLADLL 115
 + NG P + G PG PG A G G V A
 Sbjct: 126 EPSPGENGAPGQ-MGPRGLPGFPKGAAGEPGKAGERGVPPGAVGPAKDGEGAQ 184

Query: 116 QLPGAAGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
 PG A AG+RG GP A P F L G GPP A
 Sbjct: 185 GPPGPAGPAGERGE-QGP-AGSPG---FQGLPGPAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNGNGEAGGPGAARAAALPGTAGPPRPAAPPGAAPARGGPAP-GAPALPRSRGR 60
 E GE G PG R L P G GP A PG A RG P P G A A +
 Sbjct: 126 EPSPGENGAPGQMGPR---GLPGFP-GPKGAAGEPGKAGERGVPPGAVGPAKDGEGA 181

Query: 61 GRQLAERNRPRRRHRLAQPGHFGDLAAGVGRGAGGGHRRRHHVRSLLADLLQLPGA 120
 G Q P RG G PG G+ G G G+ DL PG
 Sbjct: 182 GAQGPAGPAGPAGERGEQGPAGSPG--FQGLP-GPAGPGEAGKPGEGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132
 + G+RG PG
 Sbjct: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAARAAALPGTAGPPRPAAPPGAAPARGGPAPGAPALPRSRGRQLAE 66
 GEAG G A R A PG G P P G A ARG GP PGA Q + + G A+
 Sbjct: 347 GEAGPSGPATRG---PGDR-GEPPPGPAGFA---GP-PGADGPGKAGEPGDAGAK 397

Query: 67 RNRGRPRRRHRLAQPGHFGDLAAGVGRGAGGGHRRRHHVRSLLADLLQLPGAAGAGD 126
 + P G PG G+ A +GA G G + A + PG + AG
 Sbjct: 398 GDAGPPGPAGPAGPPGPIGNVGAAPGKARGSGAGPPGATGFPGA-AGRVGPPGPSGNAG 456

Query: 127 RGLPGPDARD 137
 G PGP ++
 Sbjct: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPALPRSRGRQL 64
 G G PGA A G GP P P G A ARG P P Q PR +G
 Sbjct: 608 GPPGAPGAPGPVGPAGKSGDRGETGPAGIPGVGPAGARG---PAGP-QG-PRBKGZTG 662

Query: 65 AERNRPRRRHRLAQPGHFGDLAAGVGRGAGGGHRRRHHVRSLLA-DLLQ-LPG 119
 + + + HRG PG PG GA G RG S D L LPG
 Sbjct: 663 ZZGBRGIKHGRGFSGLQPPGPPGSPGEQGPSGASGPAGPRGPPGSAGSPGKDLNGLPG 722

Query: 120 AAGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQ 168
 G RG GP A P P P G GPP+ L +P Q
 Sbjct: 723 FIGPPGPRGRTGDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02
 Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAARAAALPGTAGPPRPAAPPGAAPARGGPAPGAP---QALPRSRGR 62
 G AG PG A R PG A GP A G A+G P P PA + P G
 Sbjct: 152 GAAGEPGKAGERGVGPPG-AVGP---AGKDGEAGQPPGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHGALAPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
Q P G + G PGDL A G G RG R + PG A
Sbjct: 208 QGLPGPAGPPGEAGKPGEQGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGPAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154
G G PG D + P G +G P
Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02
Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAARAAALPGT---AAGPFRPAAPPGAAPARG--GPA--PGAPAQLPRSQ 60
G G PG + PG A+GP P PPG G G A PG P + P +
Sbjct: 29 GPPGAPGPGFQGPGEPEPGASGPMGPRGPPGPGKNGDDGAGKPRPGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHRGALAPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G R L G P + HRG G GD +G G G + L
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAKDAGPAGPKGEPGSPGENGAPQMGPRG-LPGF 147

Query: 118 PGAAEGAGDRG 128
PG AG+ G
Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02
Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAARAAALPGT---AAGPFRPAAPPGAAPARG--GPA---PGAPAQL 55
G G PGA R A PG A G P P P G + RG GPA P PA A
Sbjct: 587 GRGSPGAKGDRGETGPAGAPGPPGAPGAPGVPVGPAGKSGDRGETGPAGIPGVPAGAR 646

Query: 56 -----PRSQGRQLAERNRPRRRHGALAPGHFGDLAAGVGRGAGGGHSRRGRHHV 108
PR +G + + + HRG G PG + +G G G
Sbjct: 647 GPAGPQGPGBKGTGZGZGBRGIKHGRGFSGLQGPFGPGSPGEQGPSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAAGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LPGP P PR AG GPP
Sbjct: 706 -----PGSAGSPGKDLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02
Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAARAAALPGTA-----AGPFRPAAPPGAAPARGGPAP-GAPAQLPRSQ 60
G G G R AA LPG AGP PG RG P G P A +
Sbjct: 287 GAPGLQGMFGERGAAGLPKPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPAGAPGDK 346

Query: 61 GRQLAERNRPRRRHGA---LAQPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G A +G P RGA +PG PG + D GA G +G
Sbjct: 347 GE--AGPSG-PAGTRGAPGDRGEPPGPPGAGFAGPPGADGQPGAKGEPGDAGAKGDAGP- 402

Query: 118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159
PG A AG G + A P+ R G G P AA R
Sbjct: 403 PGFAPGAPPGPIGNVGAPGPKGARGSGPPGATGFPGAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02
Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAARAAALPGTAA---GPPRPAAPPGAAPARGGPAP-GAPAQLPRSQ 62
+G G PGA + PG G PA PG A G P P PA ++ R + G
Sbjct: 574 SGREGAPGAEGSPGRDGSAGKGRGETGPAGAPGPPGAPGAPGVPVGPAGKSGDRGETGP 633

Query: 63 QLAERNRPRRRHGALAPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122
P RG G G+ +G G RG H R + L PG
Sbjct: 634 AGPIGPVGPAGARGPAGPQGPGB-----KGZTGZGZGBRGIKH-RGFSGLQGPFGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
G++G P A P AG RGPP +A
Sbjct: 687 SPGEQG--PS-GASGP-----ACPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02
Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPFRPAAPPGAAPARGGPA-PGAPAQLPRSQGRQLAERNRPRRRH--GALAQ 80
P G P P PG +G P PG P + P RG G P ++ G +
Sbjct: 21 PSGRPLGPPGAPGPGQGPGEPEPGASGPMGPRGPP-----GPPGKNGDDGAGK 75

Query: 81 PGHPGDAAA-GV--GRGAGGGHSRRGRHHVRSADLLQLPGAAGAGDRGH--LPGPDA 135
PG PG+ G RG G G H R + L G A AG +G PG +
Sbjct: 76 GPRPGERGPPGQARGLPGTAGLPGMKH-RGFSGLDGAKDAGPAGPKGEPGSPGENG 134

Query: 136 RDPDL-PRVFLPLAGLRGPPAAA 157
++ PR LP G GP AA

Sbjct: 135 APGQMGRG-LP--GFPKGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGGAAWARRAAALPGTAAGPPRPAAPGAAPARGGP-APGAPALPRSQGRQLA 65
GEAG G A R A G GPP PA G A G P A G P A + G

Sbjct: 347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405

Query: 66 ERNGRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHRRGR--HHVRSADLLQLPGA-- 121
P G + PG G + GA G GR A PG A

Sbjct: 406 AGPAGPPGPIGNVGAAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSNAGPPGPPGPAGK 465

Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156
EG+ G RG GP R E+ P AG +G P A

Sbjct: 466 EGSKGPRGET-GPAGRPGEVGPFGPPGAGEKGAAGA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGGAAWARRA---AALPCT--AAGPPRPAAPGAAPARGGPAPGAPAL-PRSQ 60
G G PGA R A PG A G P P P G + RG P P + P R

Sbjct: 587 GRDGSFGAKGDRGETGPAGAPGPPGAPGAPGVGPKSGDRGETGPAGPIGVGPAGAR 646

Query: 61 GRQLAERNRPRRRHAGALAQPGHPGDLA-AGVG--RGAGGGHRRGRH--HHVRSADLL 115
G A G PR +G + G G G +G G G A

Sbjct: 647 GP--AGPQG-PRBGKGTGZGZGBRGKIGHRGFSGLQGPPGPPGPSGQPSGASGPAGPR 703

Query: 116 QLPGAAGAGAGRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LPPG P PR AG GPP

Sbjct: 704 GPPGAGSPGKDGNGLPGPITG--PPGPRGRTGDAGPAGPP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGGAAWARRAAALPGTAAGPPRPAAPGAAPARGGPAPGAPALPRSQGRQ-LA 65
G G PG A + A G A P P P G A RG G P Q R +RG L

Sbjct: 485 GPPGPPGPAGEKAGAPGADGAPAGPTGP-PQGIAGQRG--VVGLPQG--RGERGFPGL 538

Query: 66 ERNGRPRRH--RGALAQPGHPGDLA---AGV---GR-GAGGGHRRGRHHVRSADLL 114
+G P + GA + G PG + AG + GR GA G GR + D

Sbjct: 539 GPSGEPGKQPSGASGERGPPGPPGPPGLAGPPGESGREGAPGAEGSPGRDGSFGAKGDR 598

Query: 115 LQL-PGAAGAGDRGHLPGP 133
+ P A G PGP

Sbjct: 599 GETGPAGAPGPPGAPGAPG 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGGAAWARRAAALPGTAAGPPRPAAPGAAPARGGPAPGAPALPRSQGRQ-LA 65
G+AG GA A + G GPP PA PG G GFA GAP R +

Sbjct: 311 GDAGPKGADGAPGKDGVRGLTGPIGPPGAPGDKGACGSPGAGTRGAPGD---RCEP 367

Query: 61 GRQLAERNRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHRRGRHHVRSADLLQLPGA 120
P G G PGD A G G G + + PG

Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPPGPIGNV----APGP 423

Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP---LAGLRGPPAAVRE 160
G G PG RV P AG GPP A +E

Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGPSNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGGAAWARRAAALPGTAAGPPRPAAPGAAPARGGPAPGAPALPRSQGRQLAE 66
G+AG PGA ++ A LG G A PG RG P A P RL

Sbjct: 275 GDAGAPGAPGSQAGPLQGMPP-GERGAAGLPGPKGDRDAGPKG-ADGAPGKDGVRGLTG 332

Query: 67 RNRGRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHRRGRHHVRSADLLQLPGAAGAGD 126
G P G PG G+ G G RG A PGA G

Sbjct: 333 PIGPP---GPAGAPGDKGAGPSGPAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154
+G PG A+ P P AG GPP

Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

ORF from 12 bp to 755 bp; peptide length: 248
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (17-39)
 LEUCINE_ZIPPER (24-46)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 3

TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802.1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.
 Length = 331

HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVLRLREVARRLERLRRRLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLVSA 89
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A
 Sbjct: 91 KIQESIEKLRALANGIEEVHRGCTISNVVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150
 Query: 90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIAATCQDQMR 132
 G+G+ A IT+ + + +S E + AT D+++
 Sbjct: 151 AGVGLGAASAVTGITTSIVEHSYSSAEAE-ASRLTATSIDRLK 193

Pedant information for DKFZphtes3_18f3, frame 2

Report for DKFZphtes3_18f3.2

{LENGTH} 193
 {MW} 19708.24
 {pI} 11.90
 {KW} All_Alpha
 {KW} LOW_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGAAMARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPQAQALPRSQ
 SEGXXX
 PRD cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccchhhhh

SEQ GRQLAERNRPRRRHRCALAQPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 PRD hhhhhhhccchhhhhhhhhcccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQFCLLHRLWLWTW
 SEGXXXXXXXXXXXXX
 PRD cchhhhhhhccccchhhhhhhhhc

SEQ LPHPGAGGGGHQ
 SEG XXXXXXXXXXXXX
 PRD ccccccccccccc

(No Prosite data available for DKFZphtes3_18f3.2)

(No Pfam data available for DKFZphtes3_18f3.2)

Pedant information for DKFZphtes3_18f3, frame 3


```

[LENGTH]      248
[MW]           27162.56
[pI]           9.92
[PROSITE]     LEUCINE_ZIPPER_2
               TRANSMEMBRANE_1
[KW]          LOW_COMPLEXITY      30.65 %
[KW]          COILED_COIL        12.10 %

SEQ            MGMRPAAAREPHGPDALRRRFQGLLLDRRGRHLRQVLRLEVARRRLERLRRLSVANVAGS
SEG           .XXXXXXXXXXXXXXXXXXXXX.XXXXXXXXXXXXXXXXXXXXXX.XXX
PRD           cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS         .
MEM           .

SEQ            SLSATGALAAIVGLSLSPTVLGTSLSLVAVGLGVATAGGAVTITSLSLIFCNSRELRVR
SEG           .XXXXXXXXX.XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD           cchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccchhhhhhh
COILS         .
MEM           .MMMMMMMMMMMMMMMMMM.

SEQ            QEIAATCQDQGREILSLSCLEFFCRWQCGRDRLQCGRNASTALYNSVVFIVFVFGSRGLFI
SEG           .
PRD           hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccc
COILS         .
MEM           .

SEQ            PRRAREGDTKVSQAVLKAKITQKLAESLESTGALDELSEQLRSRVQLCTKSRGHDLKISA
SEG           .
PRD           cccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccceeeh
COILS         .CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.
MEM           .

SEQ            DQRAGLFF
SEG           .
PRD           hhhhhccc
COILS         .
MEM           .

```

PS00029	17->39	LEUCINE_ZIPPER	PDOC00029
PS00029	24->46	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_18f3.3)

DKFZphtes3_1817

group: cell structure and motility

DKFZphtes3_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```

1 GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCCAGCAG CGGACGAGGG
51 AGGTGCCGCC GTCGCCCAGG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
101 TAAGAGATTG GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
151 AGACCTCCTG AAAAAATCCTT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
201 ACTTGTGCAG CAAAGTGGCC CAAATCCATG GCATTGTCTT AGTACCCCTG
251 AAAGGAAGCC TGTGAGCAG CATCCAGTCT ACTTGTCACT TTGAGTCCCTA
301 CATTTTGATA CCTGTGGAAG AGCATTITCA GACCTTAAAT GGAAGAGATG
351 TCTTTATTCA AGGGAACAGG ATTAATTAG GAGCTGGTTT TGCTGTCTTT
401 CTCTCAGTGC CCATTCTCTT TGAAGAAACT TTCTACAATG AAAAAGAAAG
451 GAGTTTTCAGC ATCCTGTGTA TAGCCCATCC TTGGAAGAAAG AGAGAGAGTT
501 CAGAAGAGCC TTGGGCACCC TCAGATCCCT TTCCCTGAA AACCAATTGA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTGT ACAGGAACAT
601 CGCTCTTTTC CATCGAAGAT GCTCTCTACA CCAATGCCCT CCAGCAGCTT
651 ACCACATACA CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
701 CTGAGGGACT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
751 CCTGATGAGC CAGGCAATGG AGATATACCT CCATCATGAA ATTTACAAAC
801 TGATCTTTAA ATACGTGCGG ACCATGGAGG CAAGTGAGGA TCGGGCCTTT
851 AACAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAGA GAGCTGGTCT
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTCGCAAAA
1001 CTGGTGCAGC TCATTACACA GTCTCAAGC CAGAGAGTGA ACCTGGAGAG
1051 CATGTGTGCT GATGATCTGC TATCAGTCTT GTTATACCTG CTGTGTAAGA
1101 CGGAGATCCC TAATTGGATG GCAAAATTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGCCAAAGGA TGAAGTGGGA TACTGCTGCT CCTCATTCGA
1201 AGCTGCCATT GAATATATTC GGCAAGGAAG CCTCTCTGCT AAACCCCTG
1251 AGCTGAGGGG ATTTGGAGAG AGGCTGTTCCT TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCACCCGAC TGCTGTTTA AGCACATTGC
1351 ATCAGGTAA CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTACCC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTGTAAT GATCCCTCAG TTGTCACTCC
1501 ATTTCTCAGA GACGACAGGG GGCACACCCC TCTCCATGTG CGTGTGTCT
1551 GTGGGCGAGG ATCCCTCATC GACCTCCTGG TTCCAAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCACCTGGG CCTGTGAGAA
1651 GGGCTACCAG AGCGTGACGC TGCTGTGCT GCACATCAAG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACTGGC CTGACCTAC
1751 GGCCACGAGG ACTGTGTGAA GGCTCTGTTT TACTACGACG TGGAGTCTGT
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTTGTGCA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCTCTA AGTGTGCATT
1951 AAACCTCAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCTGTGTC AGTCCCGGCA GCGTCCGCTG
2051 GACTCCATCA GCCAAGAGTC CTCCACTTCC AGCTTCTCCT CCATGTGAGC
2101 CGGCTCAAAG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGTCTGAT GGAATCTAG AAATGGTGGC TTACTTGTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGACAG
2251 AGACCCCGAA TTCTGTACCC CGTTGTGCCA GTGCCCAAG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCCTC CCGCTGCCAT GTGCCGCCCC TGACGGCCCG
2401 GCGGACCTC ATCCGCTCCT TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCCTGCCA GCAGGGCCAC
2501 TTTTCAAGTG TGAAGTGTCT GTTAGATTCT AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAACACAGC CCCTCATTTA CGCTGCTCTG GTTGGCCATC
2601 ACGAGCTTGT GGCACCTGCTG CTACAGCAGC GGGCCTCCAT TAACGCTTCT
2651 AACAAATAGG GCAACACAGC GCTGCACGAG GCTGTGATTG AAAAGCACGT

```

```

2701 CTTCTGGTGA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTCAC TGTTAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGGT TCTTTGCATG
3051 AGCCAGGAGG GCAAAGTGTC ACACCTGAGC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACAGATGCTG CGGAGACACA
3201 CGGTAGAGGA TGGCGTCGTG TCCAGGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGTAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTCAGCACC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCTT CTTCAAAGCT AATGAATACA
3501 ATTGAAACAG ACAAAAATTC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAAC CACAGCTAGA ATTCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAATGTCA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTCTTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCAT TTAGATTACA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTTCTAGAA ACATCCCTTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTACACCC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTAT TCAAAATCAA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTC CTTTTTATAA ATTTCTATAG ATTATCTGT TATTTTATG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGCTCTGAG TCCGTTTCA
4251 AATGACCTTG TGATAGGAAA ATGGTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTCTACAA AGTAATTTAT
4351 GATTTCATAA TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAATAA AAAAAAATAA
4451 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4501 G

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050
 Category: similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (945-953)

```

1 MALYDEDLK NPFYLALQKC RPDLSKVAQ IHGIVLVPCK GSLSSSIQST
51 QCFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILFEETF
101 YNEKEESFSI LCIAHPLEKR ESSEELAPS DPFSLKTIED VREFLGRHSE
151 RFDNRNIAFHF RTFRECERKS LRHHDSANA LYTKCLQOLL RDShLKMALAK
201 QEAQMNLMKQ AVEIYVHHEI YNLIFRYVGT MEASEDAAFN KITRSLQDLQ
251 OKDIGVCKPEF SFNIPRAKRE LAQLNKCTSP QOKLVCLRVQ VOLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWMN NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIROGSLSAK PPESEGFQDR LFLKQMSLL SQMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFDDC EKLVSGRIND
451 PSVVTFFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYQS VTLLLLHYKA SAEVODNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLOI GNEKGDPLH IAARWGYQGV IETLLONGAS TEIQNRLKET
601 PLKCALNSKI LSVMEAYHLS FERRQKSSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSAADPEF CHPLCQPKC APAQKRLAKV PASGLGVNVT SQDGSPLPHV
751 AALHGRADLI RLLKKGANA GARNADQAVP LHLACQGHF QVVKCLLDSN
801 AKPNKKDLGS NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFVVE LLLHGHASVQ VLNRQRQTAV DCAEQNSKIM ELLQVVPSCV

```

901 ASLDDVAETD RKEYVTVKIR KWNNSKLYDL PDEPFTROFY FVHSAGQFKG
951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRQNNLPAQS GSHAAEKNGS
1001 DWPERPGLTQ TGPGRHMLR RHTVEDAVVS QGPEAAGPLS TPQEVASARS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_1817, frame 2

TREMBL:HSU43965.1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin
G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE
ANKYRIN), N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =
380, P = 8.2e-31

>TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
ankyrin (variant 2.1)
Length = 1,719

HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31
Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
+G+T LH+AA+ GQ ++ LV+ GA VNA G TPL++A Q+ + V LL A+
Sbjct: 77 KNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEVVKFLENGAN 136

Query: 522 AEVQDNNGNTPLHLACTYGHEDCVKALVYVDVES-CRL----- 558
V +G TPL+A GHE+ V L+ Y + RL
Sbjct: 137 QNVATEDGFTPLAVALQQGHENVVHLYNGTGKGVRLPALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGDTPHLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615
D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+
Sbjct: 197 PNPDLVSKTGFTPLHIAAHYENLNVAQLLLNRGASVNFTPNQGITPLHIA--SRRGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673
L +R + E + + ++ S + G+ Q +TK +
Sbjct: 255 V-RLLDLGAQI-ETKTKDELTPHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAKRLAKVPA 732
A GD L+ VR LL++ E ++D T+ P H C R+AKV
Sbjct: 312 ---AAQGDHLDLDCVRLLLQYDAE-IDDI--TLDHLP--LHVAHC-----GHRVAKVLL 358

Query: 733 S-GLGVNVTSDQGSPLHVAALHGRADLIRLLKHGANAGARNADQAVPLHLACQGHFQ 791
G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH
Sbjct: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHL 418

Query: 792 VVKCLDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKNGNTALHEAV 851
+VK LL A PN ++ TPL A GH E+ LLQ+ A +NA T LH A
Sbjct: 419 IVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAA 478

Query: 852 IEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQSKIMELQVV 896
H +V+LLL +A+ + T + A + + +L ++
Sbjct: 479 RIGHTNMVRLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30
Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEV 524
TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +
Sbjct: 274 TPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQDHLDCVRLLLQYDAEIDD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYVDVE-----SCR----- 557
+ TPLH+A GH K L+ + +C+
Sbjct: 334 ITLDHLPPLHVAACHGHRVAKVLLDQKAPNSRALNGFTPLHIACKKNHVRVMEILLKT 393

Query: 558 ---LDIGNEKGDTPHLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSV 614
+D E G TPLH+A+ G+ ++ LLQ GAS + N ETPL A + V
Sbjct: 394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSSEAPVQSPORSVDISQESSTSSFSMSAGSRQEETKKDYREVEKLL 674
 + Y L + + + Q + P I + + A T L
 Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGH----TPLL 508

Query: 675 RAVADGDEMVRYLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734
 A + G + E V LLE ++ A T P H + K A + L +
 Sbjct: 509 IAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAARYGKVRVAELLER---D 559

Query: 735 LGVNVTSQDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVK 794
 N ++G + PLHVA H D+++LLL G + + + PLH+A +Q +V +
 Sbjct: 560 AHFNAAGKNGLTPLHVAHHNLDIVKLLPRGGSPHSPAWNGYTPHIAAKQNQVEVAR 619

Query: 795 CLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKGNLTALHEAVIEK 854
 LL N + + G TPL A GH E+VALL A+ N N G T L H E
 Sbjct: 620 SLLQYGGSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLHVASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 HV V ++L+ HG V + T + A N K+++ L
 Sbjct: 680 HVPVADVLKHGVMVDATTRMGYTPHVAHYGNIKLVKFL 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKCHPL-CFCDCEKLVSGRLNDPVSVPFSD 460
 HIAS GN V LL + + + PL C + +S L D ++
 Sbjct: 244 HIASRRGNVIMVRLLLDRGAQIETKTDELTPHCAARNGHVRSEILLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAACVQASLIDLVSKGAMVNATDYHGATPLHLACQKQYQSVTLLLLHYKA 520
 G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A
 Sbjct: 303 KNGLSPIHMAAQGDHLDVRLLLQYDAEIDDITLDHLTPLHVAACHGHHRVAKVLLDKGA 362

Query: 521 SAEVQDNNNGTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARMGYQGV 580
 + NG TPLH+AC H ++L+ +D E G TPLH+A+ G+ +
 Sbjct: 363 KPNRSALNGFTPLHIACKKNHVRVMELLK---TGASIDAVTESGLTPLHVASFMHGLPI 419

Query: 581 IETLLQNGASTEIQNRLKETPLKCAL--NSKILSVMEAYHLSFERRQKSSSEAPVQSPQR 637
 ++ LLQ GAS + N ETPL A ++++ + K + P+ R
 Sbjct: 420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ---SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDEMVRYLEWTE 693
 ++ + E++ + + AG VE +L + + +T
 Sbjct: 480 IGHNTNMVKLLLENNANPNLATTAGTTPHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query: 694 EDLEDAEDTVSAAD---PEFCHPLCQ----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741
 + V A+ HP P A L V G + +
 Sbjct: 540 LHVAARYGKVRVAELLERDAHPNAAGKNGLTPLHVAHHNLDIVKLLPRGGSPHSPA 599

Query: 742 QDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLLDSNA 801
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 600 WNGYTPHIAAKQNQVEVARSLQYGGSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKGNLTALHEAVIEKHVVFVEL 861
 N + SG TPL GH + +L+HG ++A+ G T L H A + + +V+
 Sbjct: 660 NGMLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPHVAHYGNIKLVKF 719

Query: 862 LLLHGASVQVLNK 874
 LL H A V K
 Sbjct: 720 LLHQADVNATK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQK---MCHPLCFCDCEKLVSGRLNDPVSVPFSD 458
 H+AS G+ R V LL +E + T +K H +++V +N + V +
 Sbjct: 50 HLASKEGHVVMVVELLHKEIILETTTKKGNLTALHIALAQ-DEVRELNVYGANVN--A 106

Query: 459 RDDRGTPLHVAACVQASLIDLVSKGAMVNATDYHGATPLHLACQKQYQSVTLLLLHY 518
 + +G TPL+AA ++ L+ GA N G TPL +A Q+G+++V L+Y
 Sbjct: 107 QSQKGFPTPLMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQGHENVVHLIN 166

Query: 519 KASAEVQDNNNGTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARMGY 577
 +V+ P LH+A ++D A V + D+ ++ G TPLHIA +
 Sbjct: 167 GTKGKVR-----LPALHIAAR--NODTRTAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVIELLLQNGASTEIQNRLKETPLKCAL--NSKILSVMEAYHLSFERRQKSSSEAPVQ 634
 V + LL GAS + TPL A N ++ ++ E + K P+
 Sbjct: 219 LNVAQLLNRGASVNFPTQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTDELTPHLC 278

Query: 635 PQRSVDISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDEMVRYLEWTE 693
 R+ .E + + A +TK + A GD L+ VR LL++
 Sbjct: 279 AARNGHVRSEILLDHGAPIQA-----KTKNGLSPIHM-----AAQGDHLDVRLLLQYDA 329

Query: 694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729
 E ++D D ++ C H + + P C R + +
 Sbjct: 330 E-IDDTLHDLTFLHVAHCGHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCG 788
 + +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G
 Sbjct: 389 LLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKEVETPLHMAARAG 448

Query: 789 HFQVVKLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALH 848
 H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH
 Sbjct: 449 HTEVAKYLLQNKAKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGTAPLH 508

Query: 849 EAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIM--ELL 893
 A E HV V LL AS + K+ T + A + K+ ELL
 Sbjct: 509 IAREGHVETVLALLEKEASQACMTKKGFTPLHVAKYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCPLCFDCEKLVSGRLNDPSVVTFFSRDD 461
 H+A+ G + E LL ++ H + PL + L +L P +P S
 Sbjct: 541 HVAAYGKVRVAELLERDAHPNAAGNGLTPLHVAVHNNLDIVKLLPRGSPHSPAW 600

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLHLYKAS 521
 G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+
 Sbjct: 601 NYTPLHIAAKQNOVEVARSLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 522 AEVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARMGYQGV 581
 + + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++
 Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKGV--MVDATTRMGYTPLHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRKLTPL 602
 + LLQ+ A + +L +PL
 Sbjct: 718 KFLQHQAQDVNAKTLGYSPL 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCGHFQV 793
 G VN T Q+G +PLH+A+ G ++RLLL GA + D+ PLH A + GH ++
 Sbjct: 229 GASVNFTPQNGITPLHIASRRGNVIMVRLLDGGAQIETKTKDELTPHLCAARNGHVRIS 288

Query: 794 KCLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALHEAVIE 853
 + LLD A K +G +P+ A G H + V LLLQ+ A I+ T LH A
 Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDDTLHDLTFLHVAHCG 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909
 H V ++LL GA + + LN + C + + ++MELL AS+D V E+
 Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEMLLKTG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14
 Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCPLCFDCEKLVSGRLNDPSVVTFFSRDD 461
 HIA+ G+ + V LL +E +K PL K+ L P +
 Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAKYGKVRVAELLERDAHPNAAGK 567

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLHLYKAS 521
 G TPLHVA ++ LL+ +G ++ ++G TPLH+A ++ V LL Y S
 Sbjct: 568 NGLTPLHVAVHNNLDIVKLLPRGSPHSPAWNGYTPLHIAAKQNOVEVARSLQYGG 627

Query: 522 AEVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARMGYQGV 581
 A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V
 Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLSKQANG---NLGNKSGLTPLHLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSEAPV-QSPOR 637
 + L++G + R+ TPL A N K++ + + + K +P+ Q+ Q+
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQAQDVNAKTLGYSPLHQAQ 744

Query: 638 S-VDSISQ--ESSTSSFSMSAGSQRQETTK--DYREVEKLLRAVAD 679
 D ++ ++ S S G+ K Y V +L+ V D
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTPLAIKRLGYISVTDVLKVVTD 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCGHFQV 793
 G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V
 Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684

Query: 794 KCLLDSNAKPNKKDLGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAVIE 853
 L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNAKTLGYSPLHQAAQ 744

Query: 854 KHVFFVVELLLHGASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896
 H +V LLL +GAS ++ T + A++ + ++L+VV
 Sbjct: 745 GHTDIVTLLKNGASPNVSSDGTTPLAIAKRLGYISVTDVLKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHKDTVQKMCH--PLCFCDCC-EKLVSGRLNDFSVVTFPSR 459
 H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR
 Sbjct: 310 HMAAQGDHLDVRLLLQYDAEIDDIT-LDHLTPHVAACHGHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKGYQSVTLLLLHYK 519
 G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL
 Sbjct: 368 ALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query: 520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQG 579
 AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+
 Sbjct: 428 ASPNVSNVKTETPLHMAARAGHTEVAKYLLQ---NKAKVNAKAKDDQTPHCAARIGHTN 484

Query: 580 VIETLLQNGASTEIONRLKETPLKCA 605
 +++ LL+N A+ + TPL A
 Sbjct: 485 MVKLLLENNANPNLATTAGHTPLHIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33
 Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLLDSNAK 802
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 601 NYGTPHIAAKQKQVEVARSLQYGSANAEVQGVTPHLAAQEGHAEVALLSKQAN 660

Query: 803 PNKKDLGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAVIEKHVFFVELL 862
 N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+ L
 Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 L H A V K + + A Q ++ I+ LL
 Sbjct: 721 LQHQADVNAKTLGYSPLHQAAQGHDTIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11
 Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCL 796
 + T++ G++ LH+AAL G+ +++R L+ +GAN A++ PL++A Q+ H +VVK L
 Sbjct: 71 LETTTKRGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAVIEKHV 856
 L++ A N G TPL A GH +VA L+ +G ALH A
 Sbjct: 131 LENGANQNVATEDGFTPLAVALQQGHENVVAHLINYGTK---GKVRPLALHIAARNDDT 186

Query: 857 FVVELLLHGASVQVLNKRQRTAVDCAE--QNSKIMELL 893
 +LL + + VL+K + T + A +N + +LL
 Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29
 Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKGYQSVTLLLLHYKASA 522
 GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A
 Sbjct: 503 GHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHP 562

Query: 523 EVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGVIE 582
 NG TPLH+A + + D VK L+ S N G TPLHIAA+ V
 Sbjct: 563 NAAGKNGLTPLHVAVHHNNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQKQVEVAR 619

Query: 583 TLLQNGASTEIONRLKETPLKCA 605
 +LLQ G S ++ TPL A
 Sbjct: 620 SLLQYGSANAEVQGVTPHLHA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28
 Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLL 797
 N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL
 Sbjct: 662 NLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNAPKNKKDLGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAVIEKHV 857
 A N K G +PL A GH ++V LLL++GAS N ++ G T L A ++

Sbjct: 722 QHQADVNKTKLGYSPHQAQQGHTDIVTLLKNGASPNVSSDGTTPLAIAKRLGYIS 781

Query: 858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917
 V ++L + V ++ V + S P V + DV+E + +E ++

Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827

Query: 918 KIRKK 922
 K ++

Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCV 545
 G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V

Sbjct: 35 GVDINTCNQNLGLHLASKEGHVKMVVELLHKEIILETTTKKNTALHIAALAGQDEVV 94

Query: 546 KALVYDVESCRLDIGNEKGTPLHIAARWGYQVIETLLONGASTEIQNRLKETPLKCA 605
 + LV Y ++ ++KG TPL++AA+ + V++ LL+NGA+ + TPL A

Sbjct: 95 RELVNY---GANVNAQSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVA 151

Query: 606 L 606
 L

Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06
 Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVTFFSRDDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQK 507
 L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G

Sbjct: 457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREG 514

Query: 508 YQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCVKALVYD----- 552
 + L LL +AS G TPLH+A YG + L+ D

Sbjct: 515 HVETVLALKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHPNAAGKGLTPLH 574

Query: 553 --VESCRLDI-----GNE-----KGTPLHIAARWGYQVIETLLONGASTEIQNRL 597
 V LDI G+ G TPLHIAA+ V +LLQ G S ++

Sbjct: 575 VAVHHNNLDIVKLLPRGSGPHSPAWNGYTPLHIAAKQNOVEVARSLQYGGSSANAESVQ 634

Query: 598 KETPLKCALNSKILSVMEAYHLSFERROKSSEAPVQSPQRSVDSISQESSTSSFSM-SA 656
 TPL A M A LS +Q + +S + ++QE +

Sbjct: 635 GVTPLHLAAQEGHAE-MVALLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIKH 690

Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEVMRYLLEWTEEDLEDAEDTSAADPEFCHPLCO 716
 G + T + L A G++++V++LL+ + D+ +A+ + + PL Q

Sbjct: 691 GVMVDATTR--MGYTPLVASHYGNIKLVKFLQH-QADV-NAKTKLGY-----PLHQ 740

Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSQDGSPLHVA 751
 + + + +G N S DG++PL +A

Sbjct: 741 AAQQGHTDI-VTLLKNGASPNVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07
 Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSPLHVAALHGRAOLIRLLKXGAGARNADQAVPLHLACQGHFQVVKCL 796
 V D ++ AA G D L++G + N+ LHLA ++GH ++V L

Sbjct: 5 VGFREDAATSFLLAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVKMVVEL 64

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASNKNTALHEAVIEKHV 856
 L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+

Sbjct: 65 LHKEIILETTTKKNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 124

Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885
 VV+ LL +GA+ V + T + A Q

Sbjct: 125 EVVKFLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26
 Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
 + G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y

Sbjct: 42 NQNLGLHLASKEGHVKMVVELLHKEIILETTTKKNTALHIAALAGQDEVVRELNVY 101

Query: 520 ASAEVDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGTPLHIAARWGYQG 579
 A+ Q G TPL++A H + VK L+ ++ E G TPL +A + G++

Sbjct: 102 ANVNAQSQKGFPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQOQHEN 158

Query: 580 VIETLLONGASTEIQ 594
 V+ L+ G +++

Sbjct: 159 VVAHLINYGTKGKVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21
 Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTLTLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYYDVE 554
 AT A + G ++ L H + ++ + NG LHLA GH V L++ ++
 Sbjct: 13 ATSFLLRAARSG--NLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPHLIAARWGYQGVETLLQNGASTEIQNRKETPLKCALNSKILSV 614
 L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+
 Sbjct: 71 ---LETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEV 127

Query: 615 E 615
 +
 Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01
 Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQGHFQVVKLLDSNAKPNKKDLSCNTPLIYACSGGHHELVAL 828
 + G R AD A A + G+ L + N + +G L A GH ++V
 Sbjct: 4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVVMVE 63

Query: 829 LLQHGASINASNKGNLALHEAVIEKHVVFVVELLLHGASVQLNKRQRTAVDCAEQNSK 888
 LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +
 Sbjct: 64 LLHKEIILETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 123

Query: 889 I 889
 +
 Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEEKKDYREVEKLLRAV 677
 +RRQ+ E VQ + + + Q + + Q ++ +K++R V
 Sbjct: 1614 DRRQQGQEEQVQEAQNTFTQVQGNFQNPGEQVTEEQFTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
 Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDLSGNTPLIYA 817
 +D++G T L+YA
 Sbjct: 1186 EDITGTTKLVA 1197

Pedant information for DKFZphtes3_1817, frame 2

Report for DKFZphtes3_1817.2

[LENGTH] 1050
 [MW] 117013.72
 [PI] 6.47
 [HOMOL] TREMBL:DMANKY_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,
 complete cds. 2e-45

[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13
 [FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
 3e-12

[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YDR264c] 3e-12

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08
 [FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]
 3e-08

[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
 5e-05

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
 5e-05

[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04

[BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att
 dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12
 [SCOP] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12
 [EC] phosphotransferase 1e-19
 [PIRKW] nucleus 1e-13

[PIRKB] potassium channel 5e-15
 [PIRKB] early protein 2e-13
 [PIRKB] tumor suppressor 1e-09
 [PIRKB] duplication 1e-14
 [PIRKB] tandem repeat 1e-19
 [PIRKB] heterodimer 1e-14
 [PIRKB] potassium transport 5e-15
 [PIRKB] cell cycle control 1e-10
 [PIRKB] serine/threonine-specific protein kinase 1e-19
 [PIRKB] transmembrane protein 5e-15
 [PIRKB] transport protein 5e-15
 [PIRKB] DNA binding 2e-11
 [PIRKB] oncogene 1e-08
 [PIRKB] ATP 1e-19
 [PIRKB] protein kinase inhibitor 1e-09
 [PIRKB] voltage-gated ion channel 5e-15
 [PIRKB] phosphoprotein 4e-38
 [PIRKB] apoptosis 1e-19
 [PIRKB] liver 4e-09
 [PIRKB] integrin binding 3e-16
 [PIRKB] differentiation 2e-12
 [PIRKB] transforming protein 1e-08
 [PIRKB] alternative splicing 1e-40
 [PIRKB] coiled coil 1e-14
 [PIRKB] peripheral membrane protein 2e-38
 [PIRKB] transcription factor 4e-16
 [PIRKB] transcription regulation 2e-16
 [PIRKB] nucleotide binding 5e-15
 [PIRKB] phosphoric monoester hydrolase 1e-12
 [PIRKB] cytoskeleton 8e-39
 [PIRKB] calmodulin binding 1e-19
 [PIRKB] smooth muscle 1e-12
 [SUPFAM] ankyrin 1e-40
 [SUPFAM] death-associated protein kinase 1e-19
 [SUPFAM] ankyrin repeat homology 1e-40
 [SUPFAM] protein kinase homology 1e-19
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07
 [SUPFAM] int-3 transforming protein 1e-08
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38
 [SUPFAM] notch protein 2e-12
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13
 [SUPFAM] rel homology 2e-11
 [SUPFAM] EGF homology 2e-12
 [PROSITE] ATP_GTP_A 1
 [PFAM] Ank repeat
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.05 %

SEQ MALYDELLKNPFYALQKCRPDLCCKVAQIHGIVLPCKGSLSSSIQSTCFESYILIP
 SEG
 lawCB
 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVLPILFEETFYNEKEESFSILCIAHPLEKR
 SEG
 lawCB
 SEQ ESSEELAPSDPFSLKTIEDVREFLGRHSEFDRNIASFHRTFRECEKSLRHHIDSANA
 SEG
 lawCB
 SEQ LYTCKLQQLLRDHLKMLAKQEAQMNLMKQAVEIYVHHEIYNLI FKYYVTMEASEDAAFN
 SEG
 lawCB
 SEQ KITRSLQDLQKDIGVKEPFSFNI PRAKRELAQLNKCTSPQOKLVCLRKVVQLITQSPSQ
 SEG
 lawCB
 SEQ RVNLETWCADDLLSVLLYLLVKTEIPNMNANLSYIKNFRFSSSLAKDELGYCLTSFEAAIE
 SEG
 lawCB
 SEQ YIRQGSLSAKPPESEGFQDLFLKQRMSSLSQMTSSPTDCLFKHIASGNQEVERLLSQE
 SEG
 lawCB
 SEQ DHKDTVQKMHPLCFDDEKLVSGRLNDPSVVTFFSRDDRGHTPLHVAAVCGQASLID
 SEG
 lawCB

```

SEQ      LLVSKGAMVNATDYHGATPLHLACQKGYSVTLTLLLHYKASAEVDQNNNGTPLHLACTYG
SEG
lawCB    .....
SEQ      HEDCVKALVYVDVESCRLDIGNEKGDTPLHIAARWGQCVIETLLQNGASTEIQNRKLET
SEG
lawCB    .....
SEQ      PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISOESSTSSFSSMSAGSRQ
SEG
lawCB    .....XXXXXXXXXXXXXXXXXXXXXXXXXX.....
SEQ      EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTSAADPEFCHPLCQPKC
SEG
lawCB    .....
SEQ      APAQKRLAKVPASGLGVNVTSDQGSSPLHVAALHGRADLIRLLKKHGANAGARNADQAVP
SEG
lawCB    .....CHNNHHHHHHHHHCHNNHHHHHHHHHCCCCC-CCTTTTTCCH
SEQ      LHLACQQGHFQVVKCLLDSNAKNPKKDLSGNTPLIYACSGGHHELVALLQHGHASINASN
SEG
lawCB    HHHHHHHCHHHHHHHHHHCCCTTTTCTTTTCHHHHHHHHTTHHHHHHHHCCCTTTTTEE
SEQ      NKGNTALHEAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLOQVPSCV
SEG
lawCB    TTTTEHHHHHHHHCHHHHHHHHHHCCCTTTTCTTTTCHHHHHHHHCHHHHHHHC.....
SEQ      ASLDVAETDRKEYVTVKIRKKWSKLYDLPDEPFTQFYFVHSAGQFKGKTSREIMARD
SEG
lawCB    .....
SEQ      RSVPNLTEGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPGLTQTGPHRRMLR
SEG
lawCB    .....
SEQ      RHTVEDAVVSQGPAAAGPLSTPQEVSAASRS
SEG
lawCB    .....

```

Prosite for DKF2phtes3 1817.2

PS00017 945->953 ATP GTP A PDOC00017

Pfam for DKFZphtes3 1817.2

HMM_NAME	ANK	repeat
HMM		*GyTPLHIAARYNNvEMVrILLQHGDIN* G+TPLH+AA ++ +++LL+++GA +N
Query	463	GHTPLHVAACGQASLIDLVLVSKGAMVN 490
32.12 (bits) f:	496 t:	523 Target: dkfzphtes3_1817.2 similarity to ankryns
Alignment to HMM	consensus:	
Query		*GyTPLHIAARYNNvEMVrILLQHGDIN* G TPLH+A++ + ++ LLL +A +
dkfzphtes3	496	GATPLHLACQKGYQSVTLTLLHVKASAE 523
Query	f:	529 t: 556 Target: dkfzphtes3_1817.2 similarity to ankryns
Alignment to HMM	consensus:	
HMM		*GyTPLHIAARYNNvEMVrILLQHGDIN* *G+TPLH+A+ Y+++++V+ L+ +
Query	529	GNTPLHLACTYGHEDCVKALVYYDVESC 556
42.65 (bits) f:	565 t:	592 Target: dkfzphtes3_1817.2 similarity to ankryns
Alignment to HMM	consensus:	
Query		*GyTPLHIAARYNNvEMVrILLQHGDIN* G+TPLHIAAR + +++ LLQ+GA
dkfzphtes3	565	GDPTPLHIAARWGYQGVIE7LLQNGASTE 592
Query	f:	744 t: 771 Target: dkfzphtes3_1817.2 similarity to ankryns
Alignment to HMM	consensus:	
HMM		*GyTPLHIAARYNNvEMVrILLQHGDIN* G +PLH+AA +++ +++RLL+HGA+ GSSPLHVAALHGRADLLTLLLKHGANAG
Query	744	771

36.38 (bits) f: 777 t: 804 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query *GyTPLHIAARYNNvEMVrILLQHGADIN*
PLH+A++++ ++V+ LL+ +A +N
dkfzphes3 777 QAVPLHLACQGGH:QVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM *GyTPLHIAARYNNvEMVrILLQHGADIN*
G+TPL++A+ ++ E+V LLLQNGA+IN
Query 810 GNTPLIYACSGGHHELVALLQHGASIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query *GyTPLHIAARYNNvEMVrILLQHGADIN*
G+T+LH A+++ +V +V+LLL HGA++
dkfzphes3 843 GNTALHEAVIEKHVFVVELLLHGASVQ 870

DKFZphtes3_19f19

group: testes derived

DKFZphtes3_19f19 encodes a novel 254 amino acid protein with weak similarity to S. cerevisiae protein YFL046w.

The protein contains a RGD cell attachment site.

No informative BLAST results; No predictive prosice, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp

Poly A stretch at pos. 1367, no polyadenylation signal found

```
1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTC CCAGCATTCC TGTGGTGCCA GAACACCTT GCCCGAAAAGC
101 CTGTGCGAGA TTTACCCCGT CTTCGCCCTC CCTCCACCG GAAAACTCTG
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCCGCGG CATTCTCGC CGGCCCTGGG
251 GAGAGAGTTC TTCACTACCA CAACCAAGGA GGGATATGAT AGGCCGCCAG
301 TGGATATAAC TCCTTTAGAA CAAAGGAAT TAACCTTTGA TACCCATGCA
351 TTGGTTCAGG ACTTGGAAAC TCATGGATT GACAAAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAACTG CTTTATCAAA TGTCAGCCTG GATACTATCT
451 ATAAAGAGAT GGTCACTCAA GCTCAACAGG AAATAACAGT ACAACAGCTA
501 ATGGCTCATT TGGATGCTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAAG
551 TGAATTGCA AATCTGAGAG CAGAGAATGA GAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCAACAAGTA ATGCATGAAA CCAGTCGAT CAGACGAGAT
651 AATAAAGTGG ATATCAACTT AGAAGGAGC ACAGTAACAG ATATCTTTAC
701 AGATCAAGAA AAGCACTTA TGGAAACAC TACAGAATTT ACAAAAAGCG
751 ATACTCAAAC CAAACTATT ATTTAGAGA CCAGTAATTA AATTGACGCT
801 CAAATTGCTT CCTTAAACAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTTACTTG CCGCAATA GCATTGGGAT
901 TTTTAGATT CTGCAAGTAG TATTAATGCT CATCCTGCTG TGGCTGTTGG
951 CTTCTTAGAA CACCAACCG GGAGAGATT ACTTTGAACA TTGTCAGTTG
1001 CAGCAAAAT TTAACAACA AGATTATTCG AAGTGTATAC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATAC GTGTCTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCCTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAC CATTCCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATAATTA CATTCTTCTA
1351 TAAATGAAA GATTATTACA ACAAAAAAAA AAAAAAAA AAAAA
```

BLAST Results

Entry HS419346 from database EMBL:

human STS WI-13569.

Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:

human STS SHGC-50338.

Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:

human STS WI-13893.

Score = 1578, P = 1.0e-64, identities = 358/397

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 917 bp; peptide length: 254
Category: similarity to unknown protein
Classification: no clue
Prosites motifs: RGD (15-18)

1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTTT KEGYDRRPVD
51 ITPLEQRKLT FDTHALVQDL ETHGFDKTA ETIVSALTAL SNVSLDTIYK
101 EMVTOAQOEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMIELDQ
151 VKQQLMHETS RIRADNKLDI NLESRVTDM FTDQEKQIME TTTEFTKKDT
201 QTKSISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLALALGFY
251 RFWK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19f19, frame 3

SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME
I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 138, P = 5.4e-08

>SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.
Length = 211

HSPs:

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09
Identities = 34/121 (28%), Positives = 67/121 (55%)

Query: 70 LETHGFDKTAETIVSALTALSNVSLDTIYKEMVTOAQOE-ITVQQLMAHLDIAIRKDMVI 128
LE G+ AETI + + + + +L + K + +A+QE + + QQ L IRK +
Sbjct: 46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAKQESVSFQQKRTFLQ-IRKYLET 104
Query: 129 LEKSEFANLRAENKMKIELDQVKQQLMHETSRIADNKLIDINLESRVTDMFTDQEKQL 188
+E+EF +R +K+ E++ K L + ++ +L++NLE+ R+ D T + +
Sbjct: 105 IENEFDKVRKSSDKLINEIEKTKSSLRDVKTALSEVRLNLNLEKGRMKDAATSRNTNI 164
Query: 189 ME 190
E
Sbjct: 165 HE 166

Pedant information for DKFZphtes3_19f19, frame 3

Report for DKFZphtes3_19f19.3

{LENGTH} 254
{MW} 29505.73
{pI} 6.99
{MOMOL} PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YFL046w] 8e-12
{PROSITE} RGD 1
{KW} TRANSMEMBRANE 1
{KW} LOW COMPLEXITY 5.12 %
{KW} COILED_COIL 11.02 %

SEQ MNSRQAWRLF LSQGRGDRWVSRPRGHFSPALRREFFTTTTKEGYDRRPVDITPLEQRKLT
SEG
PRD ccchhhhhhhhhccccccccccccccccchhhhhhhheeeccccccccccchhhhhhhcc
COILS
MEM
SEQ FDTHALVQDLETHGFDKTAETIVSALTALSNVSLDTIYKEMVTOAQOEITVQQLMAHLD
SEG
PRD chhhhhhhhhhhccccccccchhh
COILS

PCT/IB00/01496

Prosites for DKFZphtes3_19f19.3

PS00016 15->18 RGD PDOC00016

(No Pfam data available for DKFZphtes3_19f19.3)

DKF2phtes3_19j17

group: testes derived

DKF2phtes3_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
Poly A stretch at pos. 2740, no polyadenylation signal found

```
1 ATCTCAGCC AAATTTTTT ATTTTGGCA GAATCAGTG GCAAGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTTTTGT GTTAGTGTG AITGTTATC
101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTTTGTG TGAGAAACTT
151 TAAAGAGTAA AGCAGAAATTG AAAGTGGAAA TTTTAATTT GTAGTTCAT
201 AAAATTAAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGCAGTTTAA
251 GGTTCATAT CTTCCTCTT TTTTGGGG GCGTATGTT TTACAGGCAC
301 TTAAGTATTC ATCGAAGAGT CACCCAGTA GCGGTATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAATAAAA TGTTCCGGAG
401 ATCTGATAGT CCTGAACACA AATACAGTGA CAGCACAGT CACATGAAG
451 CCAAAATGTT GCATATCTAC AGAGTTAGAG AGAGGATGG TGGACCACT
501 TACTCTCCAC AAGAAATTC ACACAACCAC AGTCTCTTC ATAGTTCAAA
551 TTCACATTCT TCTAATCCAA GCAATAACCC AAGCAAACT TCAGATGCAC
601 CTTATGATTC TGCAGATGAC TGGTCTGAGC ATATTAGTCT TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG AACAGAAGTT TCACAATGGG AAAAACCAAA
701 AGAGTGGCTT GAAAGAGAAC AGAGACAAA AGAGCAAAAC AAGATGCCAG
751 TCAACAGCTT CCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGGAAAG GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTT GTCTCAACCA AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CAGTACAGC ACCCATCAAA ACCAGTGGT CATCCAAGT CTACCCCAAG
1001 CACTGTTCCT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTGTA TGCTAATGGA GCATCTACTT TATCAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAACAGAAA AGAAAAGAA CTACATCAGG
1151 AGACAAACCC GTATCACATT CTTGCACAA TCCTTCCAGC TCTTCTGCCT
1201 CTGGAAGTAA CCCCACATCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTCTCTCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTCTTAA CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTC TCAAGCTGCT CAGCTCTCTA
1501 CACAAGATAT CCCTCTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAAT GGGAGTGAAG AGGGTCACCT TGTCAGAAAG CTGATAAACA
1601 GCAGGAATGC CTTGTCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCTCTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCAGGCC ATCTAATCAG
1751 TCTCCAGTGT CTTTACATC TGATGCGTCA TCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACATAACAC AGCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTAGTATC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTACCA GTCAGCCACA CAGCAGCCTG TAACGTCTGA
1951 CAAGCAGCAA GGTATGAAC CTGTCTCTCC TCGAAGCTTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAAAATG CAACAGTTGT ACCACAATAT TCTTCTGCC GATCCACGTT
```



```
2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTTGTAC
2251 TGAATTAATAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATTCAGCAA
2301 CTTTGGGAGA GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCT
2351 GAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCC AATCTTAACA
2451 TTTTGGAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGATC AATTCAGGGG AAAGATACAA
2551 GATTGATTG TAAACCCCTT GAAATGTAGA TTTCTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTTG TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA TGTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA
```

BLAST Results

Entry AC005876 from database EMBLNEW:
Homo sapiens chromosome 10 clone CIT987SK-118815 map 10p11.2-10p12.1,
complete sequence.
Score = 2130, P = 0.0e+00, identities = 426/426
12 exons matching Bp 492-2740

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209
Category: questionable ORF
Classification: no clue

```
1 MSLTSDASSP RSYVSPRIST POTNTVPIKP LISTPPVSSQ PKVSTPVVKQ
51 GPVSQSATQQ PVTADKQGGH EPVSPRSLQR SSQSPSPGCP NHTSNSSNAS
101 NATVVFPQNS ARSTCSLTFA LAHFSENLI KHVQGWPAH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLRVCEIQ ATLREQRIIF LRQIKLELEK
201 LKNQNSFMV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: WW DOMAIN_1 (90-116)
WW_DOMAIN_1 (90-116)

```
1 MRDAGDPSP NKMLRRSDSP ENKYSDSTCH SKAKNVHTR VRERDGGTSY
51 SPOENSHNHS ALHSSNSHSS NPSNNSKTS DAPYDSADDW SEHISSSGKK
101 YYNCRTEVS QWEKPKEWLE REQROKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSQTs RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFDAANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPOSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAAVT
351 QASLQSIHKK FLTAGPSAFN ITSLISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSILC QKADKQOECL VWNGSIMVQR LLQPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 3

TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid
Y40B1A, N = 1, Score = 144, P = 1.8e-09

>TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A
Length = 120

HSPs:

Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09
Identities = 30/67 (44%), Positives = 43/67 (64%)

Query: 90 WSEHSSSGKKYYNCRTEVSQWEKPKW-LEREQRQKEANKMAVNSFPK---DRDYRRE 145
W+E +SSSGK YYN +TE+SQW+KP EW E +++ K VN P+ DR Y
Sbjct: 11 WTEQMSSSGKMYYNKKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query: 146 VMQATATS 153
+ Q +++S
Sbjct: 70 IGQLSSSS 77

Pedant information for DKFZphtes3_19j17, frame 2

Report for DKFZphtes3_19j17.2

[LENGTH] 209
[MW] 22873.85
[pI] 9.95
[KW] All_Alpha
[KW] LOW_COMPLEXITY 13.40 %

SEQ MSLTSDASSPRSYVSPRISTPQTNTVPIKPLISTPPVSSQPKVSTPVVKQGPVSQSATQO
SEG
PRD ccc
SEQ PVTADKQGGHEPVSPRSLQRSSQSPSPGPNHTSNSSNASNATVVPQNSSARSTCSLTPA
SEGXX
PRD ccc
SEQ LAAHFSENLIKHVQGWADHAKEQASRLREEAHNMGTIHMSEICTELKNLRLVLRVCEIQ
SEG
PRD hhhhhhchhhhhccccchhh
SEQ ATLREQRILFLRQQIKELEKLNQNSFMV
SEG
PRD hhh

(No Prosite data available for DKFZphtes3_19j17.2)

(No Pfam data available for DKFZphtes3_19j17.2)

Pedant information for DKFZphtes3_19j17, frame 3

Report for DKFZphtes3_19j17.3

[LENGTH] 436
[MW] 47716.62
[pI] 8.71
[HOMOL] TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A 6e-08
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL012w] 2e-04
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPR152c] 6e-04
[BLOCKS] BL01159 WW/rsp5/WWP domain proteins
[PROSITE] WW_DOMAIN_1 2
[PFAM] WW/rsp5/WWP domain containing proteins
[KW] All_Alpha
[KW] LOW_COMPLEXITY 22.48 %

```

SEQ MRDAGDPSPFNKMLRRSDSPENKYSdstGHsKAKNVHThRVRErDGgTSYsPQENSHNHS
SEG .....xxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ALHSSNSHSSNPSNNPSKTSdAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKEWLE
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh

SEQ REQRQKEANKMAVNSFPKDRDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQTs
SEG .....
PRD hhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccc

SEQ RHNDRDYRLPRAETHSSSTPVQHPIKPVVHPTATPSTVPSSPFTLQSDHQPKKsFDANGA
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ STLskLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPSTSSASGLNPTSAPPTSASAVPV
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ SPVPQSPiPPLLDpNLLRQLLPALQATLQLNNSNVDISKINEVLTAaVTQASLQSIHk
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccchhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhh

SEQ FLTAGPSAFNITSLISQAQLSTQDIPLHEGIQMERDThRSKWEVKGSLCQKADKQqECL
SEG .....
PRD hhccccccceehhhhhhhhhhhccccccccccccccccccccccccccccccccchhhhhhhccce

SEQ VWNGSiMVQRLLQPSG
SEG .....
PRD eccccchhhhhcccccc

```

Prosite for DKFZphtes3_19j17.3

```

PS01159 90->116 WW_DOMAIN_1 PDOC50020
PS01159 90->116 WW_DOMAIN_1 PDOC50020

```

Pfam for DKFZphtes3_19j17.3

```

HMM_NAME WW/rsp5/WWP domain containing proteins
HMM *LPsGWEeHWDpsGrpWYYWNHETkTTQWEpP*
+ ++W EH++ SG+ YY+N T+ +QWE+P
Query 86 SADDWSEHISSSGKK-YYNCRTEVSQWEKP 115

```

DKF2phtes3_lcl

group: signal transduction

DKF2phtes3_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```
1 GCGAAGTGAA GGGTGGCCCA GGTGGGCCCA GGCTGACTGA ATGTATCTCC
51 TAGCTATGGA CTAATAATA CATGGGGGGA AATAACAAG TATTCATGAG
101 GGTGAAAATG TGACCCAGCA GGAAATTAC AACTATTTTC AATTGACGTT
151 GAATAGGATG AGTCATGGAA TTTAAGTGAT TTAATGAAGA TTATACTACT
201 GGTAGATAGA AGAGCTAAAG AAAGATGGAT ACTATGATGC TGAATGTGCG
251 GAATCTGTTT GAGCAGCTTG TGCGCCGGGT GGAGATTCTC AGTGAAGGAA
301 ATGAAGTCCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCGGTAAA
351 AAGTGGCAGA GGACTGACCA TGAGCTGGGG AAATACAAGG ATCTTTTGAT
401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTAAAGCTG AAGCATGCAC
451 GTAATCAGGT GGATGTAGAG ATCAAACGGA GACAGAGAGC TGAGGCTGAC
501 TGCGAAAAGC TGGAAACGACA GATTCAAGCTG ATTCGAGAGA TGCTCATGTG
551 TGACACATCT GGCAGCATTG AACTAAGCGA GGAGCAAAA TCAGCTCTGG
601 CTTTCTCAAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAGAA
651 CTATCAACCA TTGATGAATC TGGTTCATT TTATCAGATA TCAGCTTTGA
701 CAAGACTGAT GAATCACTGG ATTGGGACTC TCCTTTGGTG AAGACTTTCA
751 AACTGAAGAA GAGAGAAAAG AGGCGCTCTA CTAGCCGACA GTTTGTTGAT
801 GGTCCTCCCTG GACTCTGAAA GAAACTCGT TCCATTGGCT CTGCAGTAGA
851 CCAGGGGAAAT GAATCCATAG TTGCAAAAC TACAGTGACT GTTCCCAATG
901 ATGGCGGGCC CATCGAAGCT GTGTCCAATA TTGAGACTGT GCCATATTGG
951 ACCAGGAGCC GAAGGAAAAC AGGTACTTTA CAACCTTGGG ACAGTGACTC
1001 CACCCTGAAC AGCAGGCAGC TGGAGCCAAG AACTGAGACA GACAGTGTGG
1051 GCACGCCACA GAGTAATGGA GGGATGCGCC TGCACTGACT TGTTCCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTTCCA TGTGGAAGC GGATAAAATT
1151 TGGCAAAATTA TCTCTGAAGT GTGAGACTGT TCGTGTGGTC TCTCATCCAG
1201 AATGTCGGGA CCGCTGTCCC CTTCCCTGCA TTCTACCCCT GATAGGAACA
1251 CCTGTCAAGA TTGAGAGGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCAATT GTTAAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCGCACA
1401 GTAAAAGAGC TGAAGAGAA ATTCTCAGA GTGAAAAC TGACCCCTCT
1451 CAGCAAGTGC GATGATATCC ATGCTATCTG TAGCCCTTCT AAAGACTTTC
1501 TTGAAAACCT CAAAGAACCT CTTCTGACCT TTCCCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAACCTG CCCAGGCCAA CAGGACACA TTAGCTTTCC
1651 TCATGATTCA CTGACAGAGA GTGGCTCAGA GTCCACATAC TAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGCCCTT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCTGCTT TCCTTGCCCT TGGAGTATTG GAGTCAGTTC
1851 ATGATGGTGG AGCAAGAGAA CATTGACCCC TCATAGTCA TTGAAAATCT
1901 AAATGCCTTT TCAACCCAC AGACACCAGA TATTAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCTGAA CATCAGCTTC TCAAGACTCC TTCACTAGT
2001 TCCCTGTAC AGAGAGTCCG TTCCACCCCT ACCAAGAACA CTCTAGATT
2051 TGGGAGCAAA AGCAAGTCTG CCACTAACCT AGGACGACAA GGCAACTTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCTGTT ACTTCCCAGC
2151 ATTGACTGAC TATAAGAAAG GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTACTCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTGATGAG GGTTTTATTA AACTATATA TATCTCCCTC TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TCTTTTGTGG GGAAGGGGGT TATTGTTCCT TTGGCTTAAA GCCAATGCT
2451 GCTCATAGAA TGATCTTTCT CTAGTTTCAT TTAGAATGAA TTTCCGTGAG
2501 ACAATGACAG AAACCTTACC TATCTGATTA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTTAGGAT
```

```

2601 GCCTCCTTCT AAGAACCAGA AGTTCTCATT CCCCATTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG COTAGCTTTG ATTGGGATGA TTAGGCTCTT AATAGTGTG
2751 AGTGGGACAA CCTTGTAAT GTGAAAGTAC AACTGCTATT TATCTCTGAT
2801 GTGGCCCTGG CTGAACCTTG GTTCACTTGG GGCTCAAAGC CAGTTTTTCT
2851 TTTAAATTGG AATTCATTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCACTGGAGC CCAACTTCTA AAGCTCAATA TATCATCCTT TGGCATCCCA
2951 ACTACAATA AAGAGTAGGC TATAGGGGAA GATTGTCAAT ATTTTGTGGT
3001 AGAAAAAGCT ACAGTCATTT TTTCTTTGCA CTTTGGATGC TGAAATTTTT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTTA
3101 AAATTATTCT TAATGCTCTG TAAAACGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCTATTG ACCCTTATCT GTAAAACACC TATTTGGGAT AATATTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAA

```

BLAST Results

Entry U82984 from database EMBLEST:
Homo sapiens DRES 56 mRNA sequence.
Score = 8775, P = 0.0e+00, identities = 1757/1758
matches 3' end

Medline entries

93074974:
Developmental regulation and neuronal expression of the mRNA of rat
n-chimaerin, a
p21rac GAP:cDNA sequence.

93024458:
A Drosophila rotund transcript expressed during spermatogenesis and
imaginal disc
morphogenesis encodes a protein which is similar to human Rac
GTPase-activating
(racGAP) proteins.

Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632
Category: similarity to known protein

```

1 MDTMMLNVRN LFEQLVRRVE ILSEGNEVQF IQLAKDFEDF RKKWQRTDHE
51 LGKYKDLMLK AETERSALDV KLKHARNQVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIQLSEE QKSALAFLNR GQPSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLOWDSS LVTFKLLKKR EKRRSTRQF VDGPPGPVKK
201 TRSIGSAVDQ GNESIVAKTT VTVPNDDGPI EAVSTIETVP YWTRSRRTKG
251 TLQFWNSDST LNSRQLEPRT ETDSVGTQPS NGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKRDCR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFVSQTSP MIPSIIVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRNLN AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPTMLQD IKRQPKVVER LLSLPLEYWS QFMVVEQENI
551 DPLHVIENSN AFSTPQTPOI KVSLLGPVTT PEHQLLKTPS SSSLSQVRVS
601 TLTKNTPRFQ SKSKSATNLG RQGNFFASPM LK

```

BLASTP hits

Entry CEK08E3.4 from database TREMBLNEW:
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit
fly (Drosophila melanogaster) (fragment)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit
fly (Drosophila melanogaster)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539_1 from database TREMBL:
 gene: "rotund"; product: "rnracGAP"; *Drosophila melanogaster* rnracGAP
 (rotund) gene, complete cds.
 Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:
 N-chimerin - rat
 Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKF2phtes3_lcl, frame 3

No Alert BLASTP hits found

Pedant information for DKF2phtes3_lcl, frame 3

Report for DKF2phtes3_lcl.3

```
[LENGTH]      632
[MW]            71026.84
[PI]            9.08
[HOMOL]         PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -
fruit fly (Drosophila melanogaster) 2e-46
[FUNCAT]        10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11
[FUNCAT]        30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]
2e-11
[FUNCAT]        03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09
[FUNCAT]        30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09
[FUNCAT]        06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09
[FUNCAT]        03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YOR127w] 5e-09
[FUNCAT]        09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08
[FUNCAT]        10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08
[BLOCKS]        BL00479B Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS]        BL00479A Phorbol esters / diacylglycerol binding domain proteins
[SCOP]          dlpbwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo
[SCOP]          dirgp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 1e-49
[PIRKW]         breakpoint cluster region 1e-19
[PIRKW]         transmembrane protein 7e-08
[PIRKW]         brain 3e-22
[PIRKW]         alternative splicing 1e-19
[PIRKW]         P-loop 2e-25
[SUPFAM]        CDC24 homology 3e-22
[SUPFAM]        bcr protein 3e-22
[SUPFAM]        myosin motor domain homology 2e-25
[SUPFAM]        pleckstrin repeat homology 4e-10
[SUPFAM]        LIM metal-binding repeat homology 2e-09
[SUPFAM]        protein kinase C zinc-binding repeat homology 5e-29
[PROSITE]       MYRISTYL 6
[PROSITE]       AMIDATION 1
[PROSITE]       CAMP_PHOSPHO_SITE 3
[PROSITE]       CK2_PHOSPHO_SITE 13
[PROSITE]       TYR_PHOSPHO_SITE 2
[PROSITE]       PKC_PHOSPHO_SITE 9
[PROSITE]       ASN_GLYCOSYLATION 1
[PROSITE]       DAG_PE_BINDING_DOMAIN 1
[PFAM]          Phorbol esters / diacylglycerol binding domain
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY 2.22 %
[KW]            COILED_COIL 8.54 %
```

```
SEQ      MDTMMLNVRNLFQQLVRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK
SEG      .....
COILS    .....CCCCCCCCCCCC
lrgp-    .....

SEQ      AETERSALDVKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSILQSEE
SEG      .....
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
lrgp-    .....

SEQ      QKSALAFLNRGQPSSSNAGNKRSLTIDESGSILSDISFDKTDSELDWSSLLVKTFLKKR
SEG      .....
COILS    .....
```

```

lrgp- .....
SEQ   EKRRSTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDGGPIEAVSTIETVP
SEG   .....
COILS .....
lrgp- .....

SEQ   YWTRSRRTGTTLQPWNSDSTLNSRQLEPRTEQSVGTQPSNGGMLRHDFVSKTVIKPESC
SEG   .....
COILS .....
lrgp- .....

SEQ   VPCGKRIKFGKLSLKCRDCRVVSHPECRDRCPLPCIPTLIGTPVKIGEGMLADFVSQTSP
SEG   .....
COILS .....
lrgp- .....

SEQ   MIPSIVVHCVNEIEQRGLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG   .....
COILS .....
lrgp- .CCHHHHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHHCCCGG-GGCCCCHHHHH

SEQ   LLKDFLRNLKEPLLTFLNRAFMEAAEITDEDNSIAAMYQAVGELPQANRDTLAFMIHL
SEG   .....
COILS .....
lrgp- HHHHHHHHTTTTTTGGGHHHHHHHTTTT-CGGGHHHHHHHHHHCCCHHHHHHHHHHHHH

SEQ   QRVAQSPTKMDVANLAKVFGPTIVAHAVPNPDVMTLQDIKRQPKVVERLLSLPLEYWS
SEG   .....
COILS .....
lrgp- HHHHHHHHHCCCHHHHHHHHGGGCC.....

SEQ   QFMVQEENIDPLHVIENSNAFSTPQTQDKVSLGPFVTTPEHQLKTPSSSSLSQQRVRS
SEG   .....
COILS .....
lrgp- .....

SEQ   TLTKNTPRFGSKSKSATNLGRQGNFFASPMK
SEG   ....
COILS .....
lrgp- .....

```

Prosites for DKFZphtes3_lcl.3

PS000001	212->216	ASN GLYCOSYLATION	PDCC000001
PS000004	141->145	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	182->186	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	246->250	CAMP_PHOSPHO_SITE	PDCC000004
PS000005	63->66	PKC_PHOSPHO_SITE	PDCC000005
PS000005	174->177	PKC_PHOSPHO_SITE	PDCC000005
PS000005	186->189	PKC_PHOSPHO_SITE	PDCC000005
PS000005	245->248	PKC_PHOSPHO_SITE	PDCC000005
PS000005	313->316	PKC_PHOSPHO_SITE	PDCC000005
PS000005	392->395	PKC_PHOSPHO_SITE	PDCC000005
PS000005	435->438	PKC_PHOSPHO_SITE	PDCC000005
PS000005	595->598	PKC_PHOSPHO_SITE	PDCC000005
PS000005	606->609	PKC_PHOSPHO_SITE	PDCC000005
PS000006	47->51	CK2_PHOSPHO_SITE	PDCC000006
PS000006	66->70	CK2_PHOSPHO_SITE	PDCC000006
PS000006	144->148	CK2_PHOSPHO_SITE	PDCC000006
PS000006	206->210	CK2_PHOSPHO_SITE	PDCC000006
PS000006	234->238	CK2_PHOSPHO_SITE	PDCC000006
PS000006	270->274	CK2_PHOSPHO_SITE	PDCC000006
PS000006	323->327	CK2_PHOSPHO_SITE	PDCC000006
PS000006	387->391	CK2_PHOSPHO_SITE	PDCC000006
PS000006	392->396	CK2_PHOSPHO_SITE	PDCC000006
PS000006	410->414	CK2_PHOSPHO_SITE	PDCC000006
PS000006	449->453	CK2_PHOSPHO_SITE	PDCC000006
PS000006	489->493	CK2_PHOSPHO_SITE	PDCC000006
PS000006	579->583	CK2_PHOSPHO_SITE	PDCC000006
PS000007	46->55	TYR_PHOSPHO_SITE	PDCC000007
PS000007	376->385	TYR_PHOSPHO_SITE	PDCC000007
PS000008	131->137	MYRISTYL	PDCC000008
PS000008	150->156	MYRISTYL	PDCC000008
PS000008	276->282	MYRISTYL	PDCC000008
PS000008	377->383	MYRISTYL	PDCC000008
PS000008	388->394	MYRISTYL	PDCC000008
PS000008	623->629	MYRISTYL	PDCC000008
PS000009	303->307	AMIDATION	PDCC000009

WO 01/12659

PCT/IB00/01496

PS00479 287->336 DAG_PE_BINDING_DOMAIN PDOC00379

Pfam for DKF2phtes3_lcl.3

HMM_NAME	Phorbol esters / diacylglycerol binding domain		
HMM	*HrFmrHTFrqPTWCDHCgeFIWGwgKQGYQCQnCGmNCHKRCHeIVPmm		
	H+E+ +T + P +C CG +I +GK ++C +C+++ H +C+ + P		
Query	287	HDFVSKTVIKPESCVPCKRI-KFGKLSLKRCRVRVSHPECRDRCPLP	334
HMM	C*		
	C		
Query	335	C	335

DKF2phtes3_lg13

group: intracellular transport and trafficking

DKF2p DKF2phtes3_lg13 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cpl51 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cpl51"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cpl51 shows haploid-specific transcription!
testis or haploid-specific transcription

Sequenced by DKF2

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```

1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
51 CATTCCTTGA ACTATTCTGC AGGTTTATATA CAAAGCTCAG AAAATACTAA
101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGCGGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTGAGG GAGTCAGAGG TGGAAATTTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAAAAAAT
401 TGCTGGTCTT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGCAGACT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCCTAGAGA AGCAGACTTC
501 CGACCTGGTT CTCTCGACCC ATCACTGCAA ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAAATCACA ACGAGAACAC AGGGGAGAAAG
601 CTCCATTTGG CGCAGGAGCA ACTCGCCTTG GCCGGGGACA AGATCCCTTC
651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAAG TCTTCCCTGA
701 GCAACATCGA GTTACTAGAA TGCCAAGTGA AGATGTTGCA GGGGGAATC
751 GGGGGATGCA TGGGTACAGA GCCTCAGAAC AAGGCTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCACAGGA TGATCTCATT
901 CAAGAATCTC GAAATAAGCT GGCCTGCAAT AACGCTTTGG TTCTGGAGCG
951 TCAAAAGGCT TTGATAAAC TACAAGCCGA TTTTGTCTCC TGTACAGCCA
1001 CCCACAGATA CCCTCTAGC TCCTCAGAAG AGTGTGAAGA CATCAAAAAG
1051 ATACTGAAGC ACTTGACGGA GCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGACTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TCGGGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCCTGCAAT GCGGCTGCA GGAGCTGCAG CTGAGATTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCTT CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCCTGCAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCAGTGCCA GGCTGTCAAC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAGAG GTCTCTCCTT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACCTCAGA TGCTGCAGAA GGAGTCTCGT ATGGCTGAGA
1751 AGGAACAACAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAATAATCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCAGAGG GAGCAAGGCT CCATCAATAG CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TCGCGCAGGA ATTTAAAAAG AAAGACAAGA CGTTGAAAGA GAATTCCAGA
2101 AAGTTGAGAG AAGAAAATGA GAATCTCCGA GCAGAGCTAC AGTGTGTGTT
2151 TACACAACTG GAATCTCTC TCAACAAATA CAACACCAGC CAGCAAGTCA
2201 TCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGCAGGCCC AGCTGGACAA AGCTCTGCAG AAGGAGAAGC ACTATCTCCA
2301 GACTACCATC ACCAAAGAAAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CCTGCCAGGA TGACCTGACA CAAGCCCTCG AGAAGCTCAA TCACGTGACC
2401 TCAGAGACAA AGAGCTGCA GCAAAGCTTG ACACAGACCC AAGAGAAGAA

```

```
2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTGC CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCGAAAAGGA
2901 CCACCTCCAC AGTGAATGG TCCACTTGCA GCAGGAAAC AAGAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAAGCCC TAGGCCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGGGGCACC TTGGGCTGGA AGGGGTGGCC CCAGGATATG GGTCAAGAAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCTCATAC
3151 TGCTAGAATC CACATCTAGC CTTGAGCAGC ATTTCCACGG GTCTTCATC
3201 AGAGGACAGT GAGTTCCAGC CCCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCAGGCT CCCAGCCTAT TTTGCAAGAC ACTAATCTTT
3301 GTTGAGTTTT GTCCACTTCC TCCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACTC TGAATAAAG ACATGAGCAT GACCAAAAAA
3401 AAAAA
```

BLAST Results

Entry AC004682 from database EMBLNEW:
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete
sequence.
Score = 1291, P = 0.0e+00, identities = 265/272

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007
Category: similarity to known protein
Prosite motifs: LEUCINE_ZIPPER (83-105)
LEUCINE_ZIPPER (90-112)
LEUCINE_ZIPPER (97-119)
LEUCINE_ZIPPER (104-126)
LEUCINE_ZIPPER (403-425)
LEUCINE_ZIPPER (410-432)
LEUCINE_ZIPPER (918-940)

```
1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQLKKKL LVLQOELEFH
101 TEELQTSYYS LRQYSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
151 ENTGELHLA QEQALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
201 MLQELGGIM GQEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVMQKVS
251 QQDDLIQELR NKLACSNALV LEREKALIKL QADFASCTAT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQKRNIMKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEST ETQKLTLLKD
401 KFLQEKDEML QELEKKLTOV QNSLLKKEKE LEKQOCHATE LEHTVKEAQ
451 DKSKEAECKA LQLEVKKLXN SLEEAQOQER LAAQQAQCK EEAALAGCHL
501 EDTQKLLQKG LLLDKQKADT IQELQRELOM LQKESMAEK EQTSNRKRVE
551 ELSLESEAL RKLENSDKEK RQLKQTVAEQ DMKMDMLDR IKHQHREQGS
601 IKCKLEEDLQ EATKILLEDK EQLKKSKEHE KLMEGELEAL QDEFKKDKT
651 LKENSRLKEE ENENLRAELO CCSTQLESSL NKYNTSQQVI QDLNKEIALQ
701 KESLMSLQAO LDKALQKEKH YLQTTITKEA YDALSRKSA CQDDLTOALE
751 KLNHVTSYTK SLQSLTOTO EKKAQLEEEI IAYEERMKKL NTELRLKRGF
801 HQESELEVHA FDKKLEEMSC QVLQWQKQHQ NDLMKLAKE EQLREFQEEM
851 AALKENLLED DKEPCCLPQW SVPKDTCLRY RGNDQIMTNL EQWAKQKVA
901 NEKLGNOLEQ QVNYIAKLSG EKDHLSVMV HLQENKKLK KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGLGWK GLPQDMGQRM DLTKYIGMPH
1001 CPGSSYC
```

BLASTP hits

Entry HS417401_1 from database TREMBL:
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete

cds.
Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862
Entry SCINTANA_1 from database TREMBL:
Saccharomyces Cerevisiae integrin analogue gene, complete cds.
Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897
Entry HS6802_2 from database TREMBL:
gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC
6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain,
ESTs, CA repeat, STS and GSS.
Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028
Entry AF092090_1 from database TREMBL:
product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

Alert BLASTP hits for DKFZphtes3_lgl3, frame 1

TREMBL:HS6802_2 product: "256 kD golgin"; H.sapiens mRNA for golgin,
N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401_1 product: "trans-Golgi p230"; Human trans-Golgi p230
mRNA, complete Cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA_1 Saccharomyces cerevisiae integrin analogue gene,
complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HS6802_2 product: "256 kD golgin"; H.sapiens mRNA for golgin
Length = 2,185

HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34
Identities = 212/816 (25%), Positives = 420/816 (51%)

Query: 145 EMGNHNEN-TGEKLHLAQELALAGOKIASLERSLNLYRDKYQSSLNIELLECQVKMLQ 203
+H + E+ G L +EQL ++ +ERSL+ YR KY ++ +L+ + K LQ
Sbjct: 119 DMDSEAEDELVGNDSLSLKEQLI---QRLRMERSLSSYRGKYSSELVTAYQMLQREKKKLQ 175
Query: 204 GELGGINGQEPENKGDHKSRIYTPCMIQEHQETQKRLSEVWQ-KVSOQDDLIQELRNK 262
G I+ Q D S RI +Q Q+ +K L E + + +D I L+ +
Sbjct: 176 G----ILSQSQ----DKSLRIAEELREELQMDQQAQKHLQEEFDASLEEKQYISVLQIQ 227
Query: 263 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKLQE 313
++ + + + K L +L+ A P S E ED K L+ LQ+
Sbjct: 228 VSLKQRLRNGPMNVDPKLPLEPQ-AEVFTKEENPESDGEPPVEDGTSVKTLETLOQ 286
Query: 314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQKRNIMKDMKLELDLHGLREETS 366
+ Q C ++ ++ L E EA+ EQ +++ K+ DLH + E+T
Sbjct: 287 RVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQELDERLQELERIK-DLH-MAEKTCL 344
Query: 367 HIERKDKDITILQCLQELQLEFTEQKLTLLKDKFLQEKDEMLQELKKLTQV--QNSL 424
+ +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q
Sbjct: 345 ITQLRADAKNLIEQLE-QDKGMVIAETKR---QMHETLEMKEEIAQLRSRIKQMTTQGE 400
Query: 425 LKKEKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKNSLEEAKQERLAAQ 484
L+++KE + + + ELE + A+ K+EA K L+AE+ + ++E+ ++ER++ Q
Sbjct: 401 LREQKE-KSERAAFELEKALSTAQ--KTEEARRK-LKAEMDEQIKTIKTSEERISLQ 456
Query: 485 QA-AQCKEEAA-LAGCHLEDTQKRLQKGLLLQKQADTIQELQRELQMLQKSSMAEKEQ 542
Q ++ K+E + E+ RLQK L +K+ A QEL ++LQ ++E E+ +
Sbjct: 457 QELSRVQEVVDVMKSSSEEQIARLQK--LHEKELARKEQLTKKLQTRERE--PQEQMK 512
Query: 543 TSNRRRVEELSLELSEALRKLKLENSDKEKRLQKT--VAEQDMKMNMDLRIKHQHQREQGS 600
+ K E L++S+ +E+ E+ +LQK + E+ K+ D+ +
Sbjct: 513 VALEKQSEY-LKISQEKQESLALAELELQKAILTESENKLRDLQEAETRYRTRILE 571
Query: 601 IKCKLEEDLQEA TKLLED----KREQLKKSKEHEKLMEG---ELEALR-QEFKKDKTL 651
++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L
Sbjct: 572 LESSLEKSLQENKQSKOLAVHLEAKNKHKEITVMVEKHKTELESKHKQDQALWTEKL 631
Query: 652 KENSRKLEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706
+ ++ + E E LR + C + E+ L +K Q I++N++ + ++ L S
Sbjct: 632 QVLKQQYQTEMEKLRK---CEQEKETLLKDKKEIFQAHEEMNEKTLEKLDVQKTELES 688
Query: 707 LQAQLDKALQKEKHYLQ--YITKEAYDALSRKSAACQDDLTQALEKLNHVTSKSLQ 764
L ++L + L K +H L+ ++ K+ D + ++ A D+ Q V S K +

Sbjct: 689 LSSELSEVL-KARHKEEELSVLKQDQTKMKQLEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRKLRGFHGESELEHVAFDCKLEEMSCQVLO 824
S+ +T+ KA L+++I E +K+ + L++ + + E ++ + +L++ S ++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLKDKHLEHQAHVENLEADIKRSEGELOQASAKLDV 802

Query: 825 WQKQHNDLKMMLAAKEEQLEFQEEAALKENLLEDDKEPCCLPQW----SVPKDT-C-R 878
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + KD C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLQKLLDLETERILLTRQVAEVEAKKDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIAKLS-GEKDLHLSVMVHLQOENK 937
L Q+ +Q KQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKIQVDLMQLEKQNSEMEQKVSILT--QV-YESKLEDGNKEQETKQILVEKENM 912

Query: 938 KLK-KEIEEKKMAENTRLCTK 958
L+ +E ++K+++ +L K

Sbjct: 913 ILQMRGQKKEIILTQKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLNS-KLL-SLQLOIKNLHDVCKRQRKTLQDN-OLCM-----EEAM 51
K+E E D E V S K L +LQ +K ++ KR ++T+Q + + C +EA+

Sbjct: 260 KEENPESDCEPVVEDGTSVKTLETLOQRVKKRQENLLKRCCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSHDKKAQALAFEESEVEFGSSKQCHLRQ----LQOLK--KKLLVLQOELFHETELQ 105
D++ + ++ + + LR ++QL+ K +++ + + H E L+

Sbjct: 320 QEQLDERLQLEKIKDLHMAEKTLITLQDLAKNLIEQLEQDKGMVIAETKQKMH-ETLE 378

Query: 106 TSYYSLRQYSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKHLAQEQL- 164
+ Q +S +++ T+ L K K + E E +T+K A+ +L

Sbjct: 379 MKEEEIAQLRSRIKQMTTQGEELREQ-KEKSERAAFELEKAL--STAQTTEEARRKLE 434

Query: 165 ALAGDKIASLERSLNLRYDKYQSSLSNI--ELLEQVVKMLQELGGIMQEPENKGDHSK 222
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 REMDEQIKTIKTSEEERISLQQLSRVQEVVDMKSSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYTPCMIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNAVLVLEREKALIKQA 282
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 QEELTKKLQTRE-REFQEQMKVALEK-SQSEYL--KISQEQEQESLALEE---LELQK 544

Query: 283 DFASCTATHRYPPSSSECECEKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAV-SE 341
A T + + +E E + + L + + +E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKRLDLQEAETYRTRILELESSLEKS---LQENKQSKOLAVHLEAKNK 600

Query: 342 QKRNIKMDMKLELDLHGLREETSABIERKDKDITI-LQCRLOELQLEFETQKTLTKKD 400
+ I + K + +L L+ + A K + + Q +++L+ E +K TL KD

Sbjct: 601 HNKEITVMVEKHKTELESILKHQDQDALWTEKLQVLKQYQTEMEKLR-EKCEQEKETLLK 659

Query: 401 K-----FLOEKDEM-LQELEKLTQVQNSLLKKEKELEKQCCMATELMTVKEAKQDKS 453
K ++E +E L++L+ K T+++ SL + E+ K + E E+V + + DK

Sbjct: 660 KEIIFQAHIEEMNEXLTLEKLDVKQTELE-SLSELSEVLKARHKEE-ELSVLKQDQTKM 717

Query: 454 K-EAECKALQAEVQKLNKSLSEAKQERLAAQQAQC-KEEAALAGCHLEDTQRKLQKGL 511
K E E K + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKDKHLEHQA 776

Query: 512 L-LDKQKADTIQELQRELQMLQKSSMAEKEQTSNRKRVEELSLELSEALRKLNSDKK 570
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGELOQASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDLETER 835

Query: 571 RQLQKTVAEQDNKMDM---LD--RIKHQHQSGSIK--CKLEEDLQEA TKLEDKREQL 623
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVEAQKDVCTELDAHKIQVDLMQLEKQNSEMEQKVSILTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMGELEALRQEFKKKDKTLKENSRLKEENENLRAELQCCSTQLESSLN 681
K +E K L+E E L+ +K K ++ ++KL + +++ + T+ ++

Sbjct: 896 NKEQEQTKQILVEKENMILQMRGQK-KEIEILTQKLSAKEDSIHILNEEYETKFRNQE 954

Query: 682 KYNTSQOVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAAC 741
K +Q +++ + + K+ L+ +A+L K L E L+ + ++ ++A + A

Sbjct: 955 KMEKVKQKAKEMQETL---KKLLDQEA LKEL--ENTALELSQKEKQFNKMLEMAQA 1009

Query: 742 QD-DLTALEKLNHVTSSETKLSQSLTQTOEKKAQLEEEIIAYEERMKKLNTLRKLRGF 800
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQKEQIE-SLTVHRR--ELNDVISIWE---KKLNQQAELQEI 1061

Query: 801 HQESELEHVAFDCKLEEMSCQVLOW--QKQHNDLKMMLAAKEEQLEFQEEAALKENLL 858
H E+++ ++++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLEKEQEVAELKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTLNELQEQ 1116

Query: 859 EDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQ--WAKQKQVANEKLGQNLREQVNYI- 915
+ L Q K L + + +L++ + ++Q V + L + + +V+ +
Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNLKSLKENTFLOEQQLVELKMLAEEDKRVSELT 1175

Query: 916 AKLSGEKDLHSMVHLQEQENKKL-KEIEEKKMKAE 951
+KL + S+ ++ NK L+ K +E KK+ E
Sbjct: 1176 SKLKTDEEFQSLKSSHEKSNKSLKEDKSLFKLSEE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26
Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLSKLLSLQDLKLNLDVCKRQKTLQDNQLCMEAMNSSHDKKQAALAFEESE 69
+E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +
Sbjct: 560 QEAETRYTRILELESSLEKSLQENKNQSKDLAVHL---EAEKNKHNEIT--VMVEKH 613

Query: 70 VEPGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQSIKELQTSDLVLLH 129
E S K H +Q +KL VL+Q+ + E+L+ Q + L K + ++
Sbjct: 614 TELESK--H-QQDALWTEKLVKQYQTEMEKLEK---CEQEKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHNTGKEL---HLAQEQALAGDKIASLERSLNLYRD 183
H + + E + + + + E+ + E L H +E+L+ + D+ + + L K D
Sbjct: 667 HIEEMNEKTLEKLDVKQTELESLSSESEVLKARHKEELSVLKQDTDKMKELEAKMD 726

Query: 184 K---YQSSLSNIELLECQVKMLQGE--LGGINGQEPENKGDHSHVRIYTPCMIQEHQE 237
+ + Q + + I + E + V + + E L + Q + K + + + +
Sbjct: 727 EQKNHHQQVDSI-IKEHEVSIQTEKALKDQINQLELLKERDK-HLKEHQAHVENLEA 784

Query: 238 TQKRLSEVMQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297
KR O+ S + D+ Q ++ ++ E+ L+LQ T R
Sbjct: 785 DIKRSEGLQQAASAKLDVFSYQS---ATHEQTAYEEQLAQQLLDLE-TERIL-- 837

Query: 298 SEECEDIKKILKHLQEQKDSOCLHVEEYQNLVKDLRVELEAVSEQKRNIMKMMKL-ELD 356
+ K + ++ OK C ++ ++ V+DL +LE + + +K + ++ E
Sbjct: 838 -----LTKQVAEVEAQKDVCTELDAHKIQVDLMQOLEKQNSEMEQVKSITQVYESK 891

Query: 357 LH-GLREETSAPIERKDKDITILQCR-LEQLQLETTQKLTLLKDKF--LQEKDEM-LQ 411
L G +E+ +K+ ILQ R Q+ ++E TQKL+ K+D L E+ E +
Sbjct: 892 LEDGNKEQEQTKQILVEKENMILQMRGQKKEIIL-TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKLTQVQNSLLK-----KEKELEKQCMATELEMTVKEAQDKSKEAECKALQAEVQ 466
EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q
Sbjct: 951 NQEKMKVQKAKEMQETLKKLLDQEAELKKELENTALELSQ-KEKQFNAKMLEM-AQ 1008

Query: 467 KLNLSLEAKQOERLAAQQAQCKEEAALAGCHLEDTRKRLQKGLLDKQKADTIQELQR 526
+A RL Q Q + + L D +K L Q+A+ +QE+
Sbjct: 1009 ANSAGISDAVS--RLETNQKEQIESLTVHRELNDVISIWEKKL---NQAEELQEIH- 1062

Query: 527 ELQMLQESSMAEKEQT-----SNRKR--EELSLESEALRKLNSDKKQRLQ 574
E+O+ +KE +AE +Q K + +E ++ L+L+ K+K
Sbjct: 1063 EIQLQKEQEVAELKQKILLFGCEKEEMKEITWLKEGVKQDTTLNELQEQKQSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHQHREQGSIKCKLEEDLQEAATKLEEDKREQLKKSKEHEKIME 634
++A+ + K+ L++++ + L+E L E L E+ + + + K +
Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNLKSLKENTFLOEQQLVELKMLAEEDKRVSELTSLKLTDD 1182

Query: 635 GELEALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694
E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +
Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLFKLSEELAIQLDICKKTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLQALDKALQKEKHYLTITKEAYDALSRKSAACQDDLT---QALE 750
K A+ + Q + K KE ++T E +A R+ Q+ L QA
Sbjct: 1242 KTNAILSR-ISHQHRTTKV--KEALLIKTCTVSEL-EAQLRQLTEEQNTLNISFQOATH 1297

Query: 751 KLNHVITSETKSLQSLTQTEKKAQLEEEIAYEERMKKLN---TELK--LRGFHQESE 805
+L ++ KS++ + Q +K L++E ++ + T+L+K
Sbjct: 1298 QLEEKENQIKSMKADIESLVTEKALQREGGQQQAASEKESCITQLKKESENINAVTL 1357

Query: 806 LEVHAFDKKLE--EMSCQVLOWQKQHQNDLKMALAAKEEQLEFQEEAALKENLLEDDKE 863
++ +KK+E +S Q+ Q QN + L+ KE + + + + K LL D +
Sbjct: 1358 MKEELKEKKVEISSLSKQLTDLNVQLQNSIS-LSEKEAIISSLRKQYDEEKCELL-DQVQ 1415

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQNAKQKQVANEKLGQNLRE---QVNYIAKLSG 920
++ K+ + D +W K+ + + N ++E Q+ +K +
Sbjct: 1416 DLSFKVDTLSKEKISALEQVDQSNKFSEWKKKQSRFTQHONTVKELOIQLELKSKEAY 1475

Query: 921 EKDH-LHSVMVHLQEQENKK---LKKEIEEKKMKAE 951
EKD ++ + L Q+NK+ LK E+E+ K K E
Sbjct: 1476 EKDEQINLLKEELDQONKRFDCLEGGEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25
Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCHEEAMNS----SHD 56
 Sbjct: 470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLTREREFEQMKVALEKSQSEYLKISQE 528

Query: 57 KKQAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQS 116
 Sbjct: 529 KEQESLALAELELQ---KKAILTESEN---KLRDLQOEAEYTRILELESSLEKSLQ 581

Query: 117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE----EMGNHNENT--GEKLHLAQEQLALA 167
 Sbjct: 582 ENKNQSKDLAVHLEAEKNKHKEITVMVEKHKTELESKHOQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNLYRDK--YQSSLS--NIELLECQVKMLQOELGGINGQEPENKGDH 220
 Sbjct: 642 MEKLREKCEQEKETLLKDKIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCHIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNALVLEKALIKL 280
 Sbjct: 701 HKLEELS--VLKD--QTDKMKQLEAKMDEQKNHQQQVDSIIKEHVSIIQTEKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338
 Sbjct: 757 QINQLELLKERDKHLKEHQAHVENLEADIKRSEGEQQASAKLDVFSQVQATHEQTKA 816

Query: 339 VSEQKRNMKMDMKLELDLHGLREETSABIERKOKDITILQCRLEQLQLETFETQKLT 398
 Sbjct: 817 YEEQLAQLQKLLDLETERILLTKQV-AEVEAQKDV---CT--ELDAHKIQVQDLMQQ 869

Query: 399 KDKFLQEKDEMLQLEKKTQVQNSLLKK-EKELEKQOCMATELEMTVKEAQDKSKEAE 457
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGKQEQOTKQILVEKENMILQMRGQKKEIE 925

Query: 458 C--KALQAEVQKLKNSLEEAQKQERLAAQQAQCKEEAALAGCHLEDTQRK--LQKGLL 513
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFKNKQKMEKVKQKAK---EMQETLKKLLDQEAAL 981

Query: 514 DRQKADTIQEL-QRELQMLQESSMAEKEQTSNRKRVEELSLESEALRKLNSDKERQ 572
 Sbjct: 982 KKELENTALELSQKEQFNQAKMLEMAQANSAGISDAVSKLETNQEIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDKMMMDLRIKHQHREQGSIRCKLEEDLOEATKLLLEDKREOLKKS---KE 628
 Sbjct: 1040 LNDVISIWEKKLNQQAELQEIHEIQLEKEQEVAEKQKILLFGCEKEEMKEITWLKE 1099

Query: 629 HEKLMEGELEALRQEFKKDKTLKENSRLKEENENLRAELQCCSTOLESSLNKYNTSQ 688
 Sbjct: 1100 EGVKQDTTLNELQEQKSAHV--NS--LAQDETKLKAHLEKLEVDNLKSLKENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLQAQL---DKALQ--KEKHYLQTTITKEA---YDALSRKSAA 740
 Sbjct: 1156 QLVELKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNKSLKEDKSLFKKLSSE-LA 1214

Query: 741 CQDDL----TQAL-----EKLNVHTSETKSLQOQSLTQTOEKKAQLEEEIIAYEERMKKL 790
 Sbjct: 1215 IQLDICCKTEALLEAKTNELINISSKTNAILSRISHCQHRHTTKVKEALLIKTCTVSEL 1274

Query: 791 NTELRLRGFHWQSELEVHAFDKKLEEMSCQVLOWQKQHNDLKMALAAKEEQLEFQEM 850
 Sbjct: 1275 EAQLROLTEEQNTLNISFQATHQLEKENQI---KSMKADIESLVTEKEAL---QREG 1327

Query: 851 AALKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGOLAE 910
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE---NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKLLKKEIEKKMAE 951
 Sbjct: 1379 LNVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFEESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLROYOSIL 118
 Sbjct: 165 QMLQREKKKLQGISQSKSLRRIAELEELQMDQAKKHLQEEFDASLEEKQOYISVL 224

Query: 119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT--GEKL---HLAQEQLALA 167
 Sbjct: 225 QTVQSLKRLRNGPMNVDLKPLQLEPQAEVFTKEENPESDGEVVDGTSVKLTETL 284

Query: 168 GDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQOELGGINGQEPENKGDHDKVRIYT 227

Sbjct: 285 ++ E L ++ QS LL ++ LQ +L + QE E D ++ 340
Query: 228 SPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASC 287
+I + ++ + ++ Q +I E + ++ L ++ E+ + +L++
Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKRQM--HETLEMKEEE-IAQLRSRIKQM 394
Query: 288 TATH---RYPPSSSEEC--EDIKKILKHLQEQKDSQCLHVEEYQNLVKDL-----RVE 335
T R SE E+++K L Q+ +++++ E +K + R+
Sbjct: 395 TTQGEELREQEKSERAFAEELEKALSTAQKTEARRKLKAEMDEQIKTIEKTSEERIS 454
Query: 336 LEA-VSEQKRNIMKDMML--ELDLHGLREETSABIERKDKDITILQCRLOLEQLETTET 392
L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E
Sbjct: 455 LQQLSRVQEVV-DVMKSSSEEQIAKLQKLHEKEKLARKEQELTK---KLQTREREFQEQ 510
Query: 393 QKLTLLKDKFLQEKDEMLQLEKKLTQVQNSLLKKELEKQCCMATELEMTVKEAQDK 452
K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+
Sbjct: 511 MKVALEKSQ--SEYKISQEKQ-----QESLAELEELQKKAIL--TESENKLRDLQEQ- 561
Query: 453 SKEAECKALQAEVQKIKNSLEEAQOER-----LAAQQAQCKEEAALAGCHLEDTOR-K 506
++ + L+ E L+ SL+E K Q + L ++ KE + H + + K
Sbjct: 562 AETRYRTRILELE--SSLEKSLQENKQSKDLAVHLEAKNKHNEITVMVEKHKTELESK 620
Query: 507 LQKGLLLDKQADTIQELQREQLQKESMAEKEQTSNRKRVESLELSEALRK--LEN 565
Q+ L ++ Q+ Q E++ L +E EKE K + + E K LE
Sbjct: 621 HQQDALWTEKLQVLKQOYQTEMEKL-REKCEQEKETLLKDKKII--FOAHIEEMNEKTEK 678
Query: 566 SDKEKRLQKTVAEQDMKMDMLDRIKHQHQEGSI-KCKLEEDLQEA-TKLEDDKR--E 621
D ++ +L+ +E ++++L + +H+ E+ S+ K + ++ QE K+ E K +
Sbjct: 679 LDVKQTELESLSSE---LSEVL-KARKMLEELSULKDQTDKMKQLEAKMDEQKNHHQ 733
Query: 622 QLKKS--KEHEKLMEGELEALRQEFKKDKDKTKENSRKLEEN---ENLRAELQCCSTQL 676
Q S KEHE ++ +AL+ + + + LKE + L+E ENL A+++ +L
Sbjct: 734 QQVDSIIKEHEVSIQRTKALKDQINQLELLKDKDKLKEHQAHVENLEADIKRSEGE 793
Query: 677 ESSLNKYNTSQVIQDLNKEALQKESLMSLQAQLDKALQKEKHYLTQTITKEAYDALSR 736
++ K + Q +++ +E L LQ +L L+ E+ L TK+ + ++
Sbjct: 794 QQASAKLDVFQSYQSATHEQTKAYEEQALQQLK-LDLETERILL---TKQVAEVEAQ 848
Query: 737 KSAACQD-----DLTQALEKLNHVTSKSLQOQSLTQTEKKAQ--LEEEIIAYEE 785
K C + DL Q LEK N SE + +SLTQ E K + +E+ +
Sbjct: 849 KKDVCTELDAHKIQVQDLNQLEKQ---SEMEQKVKSLTQVYESKLEDGKQEQETKQI 905
Query: 786 RMKLNTELRKLRGFHQESELVHAFDKKLEEMSCQVL--QWQKHQNDLKLMAKEEQ 843
+K N L+ G Q+ E+E+ +E S +L +++ +N K + +++
Sbjct: 906 LVEKENMILQMHG--QKKEIILTQKLSAKEDSIHILNEEYETKFKNQEKMEKVKQKA 963
Query: 844 RFQOEEMAALKENLEDDKEPCCLPQWSVPKDTCLRYGNDQIMTNLEQWAKQKV--- 899
+E QE LK+ LL+ ++ L + + L + Q ++ +A+
Sbjct: 964 KEMQE---TLKKLLDQEA---LKK-ELENTALELSQKEKQFNKMLEMAQANSAGISD 1016
Query: 900 ANEKLGNLREQVNYIAKLSG-EKDHLHSMVH-LQENKKLKK--EIEEKKMAENTRL 955
A +L +EQ+ + ++ E + + S+ L Q+ ++L++ EI+ ++ + E L
Sbjct: 1017 AVSRLETNQKEQIESLTVHRRELNDVISIWEKKLNQQAEELOEIHEIQLEKEQEVAEL 1076
Query: 956 CTKALGPSRTSTQREKVCGLGWKGLPQD 985
K L E + K L +G+ QD
Sbjct: 1077 KQKIL-LFGCEKEEMNKETLWKEEGVKQD 1105
Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25
Identities = 220/907 (24%), Positives = 444/907 (48%)
Query: 67 ESEVEFGSSKQCHLRLOQLKKLLVLQLEFHTTEELQTSYSLRQYQSILE---KOTS 123
E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+
Sbjct: 123 EAEDLVGNSDSLNEQLIQLRRMERSLSSYRGKYSLVATYQMLQREKKKLGILSQSQ 182
Query: 124 DLVLLHHKCLKEDEVILYEEEMGNHNTGEKHLHAEQLALAGDKIASLERSLNLVYRD 183
D L +L+E+ + +++ H + E+ + E+ I+ L+ +L+ +
Sbjct: 183 DKSL-RRIAELREE--LQMDQAKKHLQ---EEDFASLEE---KDQYISVLQTVSLLKQ 233
Query: 184 KYQSSLSNIELLECQVKMLQELGGINGQE-PENKG----DHKSVR-IYTPCMIQEHQ 236
++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +
Sbjct: 234 RLNGPMNVVVLK-PLPQLEPQAEVFTKEENPESDGEPPVEDGTSVKTLETQQRVKQE 292
Query: 237 ETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHRYPPS 296
ER E Q +Q L+ KA L ER + L K+ D T
Sbjct: 293 NLLKRCCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTCLIT-- 346
Query: 297 SSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMLKELD 356
+ D K +++ L++ K + E + + + L ++ E ++ Q R+ +K M +
Sbjct: 347 ---QLRDAKNLIEQLEQDKGM---VIAETKRQMHEHTEMKEEEIA-QLRSRIKQMTTQGE 400

Query: 357 LHGLREETS-AHIERKDKDITILQCRLOE----LQLEFTETQKLTLLKKDFLQEKDEMLQ 411
 L +E++ A E +K ++ Q + +E L+ E E K T++K +E+ + Q
 Sbjct: 401 LREQEKESERAAFELEKALSTAQ-KTEEARRLKAEMDEQIK-TIEKTSE-EERISLQQ 457

Query: 412 ELEKKLTQVQNSLLKK-EKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKN 470
 EL + +V + + K E+++ K Q + E E+ KE Q+ +K+ + + + Q +K
 Sbjct: 458 ELSRVKQEVVDVNMKSSSEQIAKLQKLH-EKELARKE--QELTKLQTREREFQEQ-MKV 513

Query: 471 SLEEAQQERLAAQQAQCKEEAALAGCHLEDQTKLQ-KGLLD-KQKADTTIQELQREL 528
 +LE++ Q E L Q + +E AL L+ + + L D +Q+A+T + EL
 Sbjct: 514 ALEKS-QSEYLKISQEKQESLAELELQKKAILTESENKRLQDQEAETRYRTRILEL 572

Query: 529 QMLQKSSMAEKEQTSNRKRVEELSLELSEALRKLENS-DKEKRLQKTVAEQDMKMDM 587
 + E S+ E + S V L E ++ + + +K K +L+ +QD +
 Sbjct: 573 ES-SLEKSLQENKNQSKDLAVH-LEAEKNKNKEITVMVEKHKTELESKHHQDQALWTEK 630

Query: 588 LDRIKHQHR-EQGSICKLEEDLQEAATKLEEDKRE--QLKKSKEHEKLMGELEALROEF 644
 L +K Q++ E ++ K E QE LL+DK Q + +EK +E +L+ + E
 Sbjct: 631 LQVLKQYQTEMEKREKCE--QEKETLLKQKEIIFQAHIEEMNKTELE-KLDVKQTEL 686

Query: 645 KKKDKTLKE--NSR-KLEENENLRAELQCCSTQLESSLNKY-NTSQVQIDLNKE--IA 698
 + L E +R KLEEE L+ + +LE+ + + N QQ + + KE ++
 Sbjct: 687 ESLSELSEVLKARHKLEELSVLKDQTKMKQLEAKMDEQKHHQVQVDSIIEHEVS 746

Query: 699 LQK-ESLMSLQA-QLDKAL-QKEKHYLTQTTKEAYDALSRSK-----AACQDDLTQAL 749
 +Q+ E + Q QL+ L +++KH + E +A ++S A+ + D+ Q+
 Sbjct: 747 IQRTKALKDQINQLELLKDKRDKLKEHQAHVENLEADIKRSEGLQQAQAKLDVFSY 806

Query: 750 EKLNHVTSETKSLQSSLTQTEKRAQLEEEIAYEERMKKLTELRLKRGFHOESELEVH 809
 + H +TK+ ++ L Q Q+K LE E I +++ + + + + + + +
 Sbjct: 807 QSATH--EQTKAYEEQLAQQLKLLDLETERILLTKQVAEVEAQQKDVCTELDAHKIQVQ 864

Query: 810 AFDKKEEMSCQVLQWQKHQN--DLKMLAAKEQLREFQEEMAALKENLL----EDDKE 863
 ++LE+ + + Q K + K+ +EQ E + + + KEN+ + E K+
 Sbjct: 865 DLMQQLERQNSEMEQKVKSLTQVYESKLEEDGNKEQ--EQTKQILVEKENHILQMRQEQK 922

Query: 864 PC-CLPO-WSVPKDTCLYRGNDQIMTNLE-QWAKQKQVANE--KLGNLQREOV-NYIAK 917
 L Q S +D+ + N++ T + Q K +KV + ++ L++++ + AK
 Sbjct: 923 EIEILTQKLSAKEDSIHIL--NEEYETKFKNQEKMEKVKQKAKEMQETLKKLLDQEAQ 980

Query: 918 LSGEKDHLHSMVHLQEQENKLLKEIEEKKMAENTRLCTKALGPSRTESTOREKV 973
 L K L + + L Q+ K+ ++ E M N+ + A+ SR E+ Q+E++
 Sbjct: 981 L---KKELENTALELSQKEKQFNAKMLE--MAQANSAGISDAV--SRLETNQEQI 1029

Score = 318 (47.7 bits), Expect = 4.4e-24, P = 4.4e-24
 Identities = 184/827 (22%), Positives = 405/827 (48%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQKRLQDNQLCMEAMNSSHDKK-Q 59
 ++ E G + + S S + L+ ++ + ++ L++ ++ + D Q
 Sbjct: 1323 LQKEGGNQQAASEKESCITQLKKELSENINAVTLMKEELKEKKEVSISSLSKQLTDLNVQ 1382

Query: 60 AQ-ALAFEESEVEFGSSKQCHLRQLQQLKLLVLQOELEFHTTEELQTSYYS-LRQYQS- 116
 Q +++ E E S + +Q + K +LL Q+L F + L S L Q
 Sbjct: 1383 LQNSISLSEKAAISSLR---KQYDEKCELLDQVQDLSFKVDTLSEKISALEQVDDW 1438

Query: 117 ---ILE-KQTSOLVLLHHCKLKEDEVILYEEEMGNHNTGEXKLHLAQELALAGDKIA 172
 E K+ + H +KE ++ L + + + + E++L +E+L +
 Sbjct: 1439 SNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAYEKD--EQINLLKEELDQQNKRF 1496

Query: 173 SLERSLNLYRDKYQSSLSNIEL-LEQVXMLQGLGGIMQEP-ENKGDHKSRIYTSPC 230
 L+ + + K + SN+E L+ Q + EL + Q+ E + + + + Y
 Sbjct: 1497 CLKGEMEDDKSKMEKKESNLETLSQTARIM-ELEDHITQKTIEISLNEVLKNYNQK 1555

Query: 231 MIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKLQADFASCTAT 290
 I EH+E ++L + ++D+ ++E K+ L LE + +K + +
 Sbjct: 1556 DI-EHKELVQKLQHFQELGEEKDNRVKEAEEKI----LTLENQVYSMAELETKKKELE 1609

Query: 291 HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVE-EYQNLVKDLRVELEAVSEQKRNIMKD 349
 H S+E E+K + L+ + + + + + + + + + L+ +E+K + +
 Sbjct: 1610 HVNLSVKSKE-EELKALEDRLESESAKLAELKRKAQKIAIAIKQLLSQMEER---EE 1664

Query: 350 MMKLELDLHGLREETS-AHIERKDKDITILQCRLOELQLEFTETQKL--TLKKDKFLQEKD 407
 K + H E L + + + + + + IL+ +L+ + +ET + + K E++
 Sbjct: 1665 QYKKGTEH--LSELNTKLQEREREVHILEEKLKSVESQSETLIVPSAKNVAAYTEQE 1722

Query: 408 EM-----LQEL-EKKLTQVQNSLLKKEKEL-----EKQCCMATELEMTVK-EAKQDKSKE 455
 E +Q+ E+K++ +Q +L +KEK L EK++ + + EM + + + K +
 Sbjct: 1723 EADSQGCQKTYEEKISVLQRNLTEKRLQRVGQKEETVSHFEMRCQYERLIKLEH 1782

Query: 456 AECKAL--QAEVQKLKNSLEEAQQERLAAQQAQCK--EEAALAGCHLEDQTKRLQKGL 511

Sbjct: 1783 AEK Q+ + L+ LEE ++ L Q + + + A +LE+ +QK L
Query: 512 LLDKQKADTIQELQRELOMLQKESMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569
++K T Q L++++ L +S + +++ +R +EEL+ E +AL++++ +K
Sbjct: 1843 --QEKELTCQILEQKIKEL--DSCLVRQKEV-HRVEEELTSRYEKILQALQMDGRNKP 1896
Query: 570 KRQLOKTVAEQD---MKMNDMLDRIKHQHREQSGIKCKLEEDLQEA TKLLEDKREQLKK- 625
L++ E+ + +L ++ QH + E + Q+ K + ++ L+
Sbjct: 1897 TELLEENTEESKSHLVQPKLLSNMEAQHNDLEFLAGAEERKQLGKEIVRLQKDLRML 1956
Query: 626 SKEHEKIMEGEALRQEFKKKDKTLKENSRLKEENENLRAELQCCSTQLESSLNKYNT 685
KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT
Sbjct: 1957 RKEHQ---ELEILKKEYDQ-----EREKIKQEDEL--ELKHNT-LQLMREFNT 2003
Query: 686 S-QQVIQDLNKEIALQKESLMSLQALQKALQKEKHYLQTTITKEAYDALSRKSAACQDD 744
Q Q+L I ++A+L ++ Q+E + L I E D L R +A ++
Sbjct: 2004 QLAQKEQELMTIKETINKAQEVAELLESHEETNQLLKIA-EXDDDLKR-TAKRYEE 2061
Query: 745 LTQALEKLNHVTSKSLQOQSLTQTEKKAQ-LEELIAYEERMK--KLNTLRLKRGFH 801
+ A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +
Sbjct: 2062 ILDAREE--EMTAKVRDLQTEELQKKYQKLEQENPGDNVTIMELOTQLAQKTTLI 2119
Query: 802 QESELEVHAFDKKLEEMSCQVLOWQK 827
+S+L+ F +++ + ++ ++K
Sbjct: 2120 SDSKLKEQEFREQIHNLEDRKKYK 2145
Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24
Identities = 213/977 (21%), Positives = 454/977 (46%)
Query: 4 EAGERD-REVSSLSKLLSLQLD-INKLHDVCKRQRTLQDNQLCMEAMNSSHDKKQAO 61
E R+ +V S+ K L+ Q + ++ +H++ + Q K + +L + + ++ +
Sbjct: 1034 EVHRRELDNDVISIWEKKLNQQAELQEIHEI-QLQKEQEVALLQKILLFGCEKEEMNK 1092
Query: 62 ALAFEESEVEFGSSKQCHLRQLO-QLKKLL---VLQOE--LEFTEELQTSYSLRQY 114
+ + + E G + L +LQ QK+K + Q E L+ H E+L+ +
Sbjct: 1093 EITWLKEE---GVKQDITLNLQELQKQSAHVNSLAQDETCLKAHLEKLEVDLNLKSLKE 1149
Query: 115 QSILEKQTSDLVLLHHHCKLEDEV---ILYEEEMGNHNTGKHLHQAQELLAGDKI 171
+ L++Q +L +L K K E+ + +E +++ EK + +E +L K+
Sbjct: 1150 NTFLEQLVLEKMLAEEDKRRVSELTSLKLTTOEEFQSLKSSHEKSNKSLKEDKSLFKRL 1209
Query: 172 AS-LERSLNYRDKYQSSLS--NIELLECQVKMLQELGGIMQEPENKGOHKSVRITYS 228
+ L L++ K ++ L EL+ L I +++ K +
Sbjct: 1210 SEELAIQLDICCKTEALLEAKTNELINISSKTNAILSRI--SHCOHRTTKVKEALLIK 1267
Query: 229 PCMIQEHQ-----ETQKRLSEVQKVSQO-DDLIQELRNKLACSNALVLEREALIKL 280
C + E + E Q L+ +Q+ + Q ++ +++ A +LV E+E L
Sbjct: 1268 TCTVSELAQLRQLTEEQNTLNISFQQAHLQEEKNQIKSMKADIESLVTEKEA---L 1323
Query: 281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340
Q + + + S E C I ++ K L E ++ L EE +K+ +VE+ ++S
Sbjct: 1324 QKEGNN---OQQAASEKESC--ITQLKKESENINAVTLMKEE----LKEKKVEISSLS 1373
Query: 341 EQKRNIMKDMKLELDLHGLREETSAMIERKDKDITILQCRLOEL--QLEFTETQKLT-L 397
+Q ++ + + L S+ ++ D++ L ++Q+L +++ +K++ L
Sbjct: 1374 KQLTDLNVQLQN-SISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVDTLSKERISAL 1432
Query: 398 KK-DKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQOCMATELEMTV---KEAQDKS 453
++ D + + E ++ + + TQ QN++ + +LE + A E + + KE ++
Sbjct: 1433 EQVDDWSNKFSEWKKASRFTQHMNTVKELQIQLELKSKEAYEKDEQINLLKEELDQQN 1492
Query: 454 KEAECKALQAEVQKLKNSLEEAKQERLAAQQAQKEEAALAGCHLE-DTORKLQKGLL 512
K +C + E K K +E+ + L Q A +E + +E ++ +K
Sbjct: 1493 KRFDCLEKGENEDKSKMEKKESNLETSLKQTAIMELEDHITQKTIEIESLNEVLKNY- 1551
Query: 513 LDKQKADTIQELQRELOMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKQ 572
+ +QK +EL ++LQ Q+ + +++ L ++ +LE KE
Sbjct: 1552 -NQKQDIEHKELVQKLQHFQELGEEKDNVKEAEKILTLENQVYSMAKAELETKKKELEH 1610
Query: 573 LQKTVAEQDMKNDMLDRIKHQHREQ-GSIKCKLEEDLQEA TKL---EDKREQLKSK 627
+ +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK
Sbjct: 1611 VNLVSKSKEELKALEDRLESESAKLAELKRAEQIAIAIKQLLSQMEKEEQYKGT 1670
Query: 628 EHEKIMEGEALRQEFKKKDKTLKENSRLKEE-ENENL---RAELQCCSTQLESSLNK 682
E EL QE +++ L+E + +E ++E L A+ T+ E + ++
Sbjct: 1671 ESHL---SELNTKQEREREVHILEEKLKSVSSQSETLIVPSAKNVAAYTEQEADSQ 1727
Query: 683 ---YNTSQQVIQDLNKEIALQKESLMSLQALQKALQKEKHYLQTTITKEAYDALSRKSA 739
T ++ I L + + +KE L+ Q +K H+ +E L A
Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFEMRCQYQERLIKLEHAE 1785

Query: 740 ACQDDLTQALEKLNHVTSET--KSLQSSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLK 797
+D Q++ +H+ E K++ SL Q + + + I ++ ++ + ++K
Sbjct: 1786 KQHED--QSM--IGHLQEELEKNNKYSLIVAHVEKEGGKNNIAKQNLNVFDDVQRT 1841

Query: 798 RGFHQESELEVHAFDKKLEEM-SCQVLQWQKHQNDLKMALAEQQLREFQEEMAALKEN 856
QE EL ++K++E+ SC V Q ++ H+ +++ L +K E+L+ Q+ K
Sbjct: 1842 L---QEKELTCQILEQKIKELDSCLVRQ-KEVHRVEMEELTSKYELQALQMDGRNKPT 1897

Query: 857 -LLEDDKEPCCLPQWSPKDTCLRYRGNDQIMTNLEQWAKQKQVANKEKLNQLEQVNYI 915
LLE++ E PK + ++ + L A++K +KLG ++ +
Sbjct: 1898 ELLEENTEESKSHLVQPKLLSNMEAQHNDEFLKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query: 916 AKLSGE-KDHLHSMVMHLQENK-KLKEIEEKKMAENTRLCTKALGPSRTESTOREK 972
L E + L + QE + K+K+E + ++K +T + + T+ Q+E+
Sbjct: 1954 RMLRKEHQELETILKKEYDQEREKIKQEDELKHNST--LKQLMREFNTQLAQREQ 2010

Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22
Identities = 221/952 (23%), Positives = 441/952 (46%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRORKTLDQNL---CMEEAMNSSHD- 56
+K A E R+VS L SKL + ++L ++ K+L+D L + E + D
Sbjct: 1160 LKMLAEEDKRVSELTSLKTTDEEFQSKSSHEKSNKSLDKSLEFKLSEELAIQLDI 1219

Query: 57 --KKQAQALAFESE-VEFGSSK-QCHLRQLQKLLVQLEFHT---EELQTSYY 109
KK L + +E + SSK L ++ + + ++ L T EL+
Sbjct: 1220 CCKTEALAEAKTNELINISSKTNAILSRISHCHRTTKVKEALLTKTCTVSELAQLR 1279

Query: 110 SLRQYQSILEKQTSIDLVLHHCKLKEDEVILYEEEMGNHNTGERLHLAQE---QLAL 166
L + Q+ L H + KE+++ + ++ EK L +E Q
Sbjct: 1280 QLTEEQNTLNISFQAT---HQLEEKENQIKSMKADI---ESLVTEKEALQEGGNQQA 1333

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVRLQELGGIMQEPENKGDHVKVRIY 226
A +K E + + + ++ + L++ +K + E + Q + V++
Sbjct: 1334 ASEK---ESCITQLKSELSENINAVTLMEELKEKKVEISSLSKQLTD----LNVQL 1384

Query: 227 TSPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFAS 286
S + ++ + + + + D +Q+L K+ + L E+ AL + + D+++
Sbjct: 1385 NSISLSEKAAISSLRKQYDEEKCELLDQVQDLSEK---DTLSKEKISALEQVD-DWSN 1440

Query: 287 CTATHRYPPSS--SEECEDIKKLKHQLQKQKQCLHVEEYQNLVKD-----LRVE-LE 337
+ + S ++ +K++ L E K + +E NL+K+ R + L+
Sbjct: 1441 KFSEWKKKAQSRFTQHONTVKELQIQL-ELKSKEAYEKDEQINLLKEELDQKNRFDCLK 1499

Query: 338 AVSEQKRNIN-KMMKLELDLHGLRE---ETSAHIERKDKDITILQCRQLQEL-QLETTET 392
E ++ M K LE +L E HI +K +I L L+ Q + E
Sbjct: 1500 GEMEDDKSKMEKKESNLETLSQATARIMELEDHITQKTIEISLNEVLKNYNQOKDIEH 1559

Query: 393 QKLTLLKDKFLQ--EKDEMLQELKLLTQVQNSLLKKEKELEKQOCMALENTVKEAK 449
++L K F + EKD ++E E+K+ ++N + + ELE ++ + ++VK
Sbjct: 1560 KELVQKLQHFQELGEEKDNVKEAEEKILTLENQVYSMAELETKKKELEHVNLSVK--- 1616

Query: 450 QDKSKEAECKALQAEVQKLKNSLEEAKQOERLAAQQAQCKEEAALAGCHLEDTRKRLQK 509
SKE E KAL+ ++ S + + +R A Q+ A K++ +E+ + + +K
Sbjct: 1617 ---SKEELKALEDRLES--ESAAKLAELKRAEQKIAAIKKQLL---SQMEEKEEQYK 1668

Query: 510 GLLLDKQKADT-IQELQRELQMLQKESMAKEQTSNRKRVEELSLELSEALRKLENSDK 568
G + +T +QE +RE+ +L+++ E Q+ + S + A + E +D
Sbjct: 1669 GTESHLSELNKLQEREREVHILEEKLKSVSSQSETL--IVPRSAKNVAAYTEQEADS 1726

Query: 569 E---KRQLQK-TVAEQDMKMD-MLDRKHQHQSGSIKCKLEEDLQEATKLLDKREQ 622
+ K +K +K +V +++ + +L R+ Q +E ++ + E Q +L+ K E
Sbjct: 1727 QGCVQKTYEEKISVLQRLNTEKEKLLQVRG-QEKEE-TVSSHFMRCQYQERLI--KLEH 1782

Query: 623 LKKSKEHE-KLMGEL-EALRQEFKKKDKTKENSRLKEEENENLRAELQCCSTOLESSL 680
+ +K+HE + M G L E L :+ KK + ++ K E N++A+ LE
Sbjct: 1783 AE-AQKHEDQSMIGHLOEELEKNNKYSLIVAHVEK-EGGKNNIAK-----QNLE--- 1832

Query: 681 NKYNYSQQVIQDLNKEIALQKESLMSLQAQLDKAL--QKEKHYLOTTITKEAYDALSR-K 737
N ++ Q+ +Q+ KE+ Q L +LD L OKE H ++ Y+ L +
Sbjct: 1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQ 1888

Query: 738 SAACQDDLTQALEKLNHVTSETKSLQSSLTQTQEKKAQ-LEEEIIAYEERMKKLNTL-- 794
++ T+ LE+ S++ +Q L E + LE ++ E +KL E+
Sbjct: 1889 QMDGRNKPTELLEENTEESKSHLVQPKLLSNMEAQHNDEFLKLAGAEREKQKLGKEIVR 1948

Query: 795 --RKLGRFHQSELEVHAFDKKLEEMSCQVLQWQKHQNDLKMALAEQQLREFQEEMAA 852
+ LR +E + E+ K+ ++ + ++ Q+Q +LK + ++ +REF ++A
Sbjct: 1949 LQKDLRMLRKEHQELETILKKEYDQEREKIK-QEDELKHNSTLKQLMREFNTQLAQ 2007

Query: 853 LKENLLEDDKEPCCLPQWSPKDTCLRYRGNDQIMTNLEQWAKQKQVANKEKLNQLEQV 912

Sbjct: 2008 ++ L KE Q V + + Q TN Q K K+A EK + R
KEQLEMTIKETINKAQ-EVEAEELLES----QEETN--QLLK--KIA-EKDDDLKRTAK 2057

Query: 913 NYIAKLSGKDHLSVMVHLQOENKKLKEIEKKMKAEN 952
Y L ++ + + LQ + ++L+K+ ++K + EN

Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQOKLEQEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22
Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQDIKN--LHDVCKRQRKTLQDNQLCMEAMNSSHDKK 58
+KD+ + +N K L +LD+K L + + L+ +EE ++ D+

Sbjct: 657 LKDKIIFQAHIEEMNEKTLE-KLDVQKTELESLSSELSEVLKARHK-LEEELSVLKQQT 714

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLV-LQOELEFHTTEELQTSYSLRQYQSI 117
+E E + K H +Q+ +K+ V +Q+ + ++ L++

Sbjct: 715 DKMK--QELEAKMDEQKNHHQQQVDSIIKEHEVSIQRTKALKKDQINQLELLKERDKH 771

Query: 118 LEKQTSDLVLLHHHCKLKEDEVILYEEEMG--NHNENTGEKHLHAQEQALAGDKIASL 174
L++ + + L K E E+ ++ T E+ +EOLA K+ L

Sbjct: 772 LKEHQAHVENLEADIKRSEGELQQAASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDL 831

Query: 175 ERSNLNRYDKYQSSLSNIELLECQVKMLQGEGLGIMGO-EPENKGDHVKVRIYTSPCMIQ 233
E L + + + + ++ +M Q E +N KV+ T

Sbjct: 832 ETERILLTKQVAEVAQKKDVCTELDAHKIQVQDLMQQLKQNSEMEQKVKSLTQ-VYES 890

Query: 234 EHQETQKRLSEVWQKVSQDDLIQELRN---KLACSNAVLEREKALIKLQADFASCTA 289
+ ++ K + Q + ++++I ++R ++ + +E ++ L ++ +

Sbjct: 891 KLEDGNKEQEQTQILVEKENMILQREGQKKEIILTQKLSAKEDSIHILNEEYET--- 947

Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRRNIMKD 349
++ + ++ E +K+ K +QE + L E L K+L +S++++

Sbjct: 948 --KFK-NOEKKMEKVKQAKEMQETLKKLLDQEA--KLKKELENTALELSQEKQKFNK 1002

Query: 350 MMKL-ELDLHGLREETA-HIERKDKDITILQCRLEQLQLEFTETQKTLKKDKFLQEKD 407
M+++ + + G+ + S +K++ ++ +EL + +K ++ +LQE

Sbjct: 1003 MLEMAQANSAGISDAVSRLETNQKEQIESLTVHRRLENDVISIMEKKNQAAELQEIH 1062

Query: 408 EM-LQELEKKLTQVQNSLLK---KEKELEKQCMATE----LEMTVKEAQD-KSKEAEC 458
E+ LQE E+++ ++ +L +++E+ K+ E +T+ E++ K K A

Sbjct: 1063 EIQLEKEQEVAELKQKILLFGCEKEEMKEITWLKEGVKQDTTLNELQQLKQSAHV 1122

Query: 459 KALQAEVQKLSLEAKQERLAAQAAQCKEEAALAGCHLEDTQKRLQKGLLDKQKA 518
+L + K LK LE+ + +E+ + + +L K K

Sbjct: 1123 NSLAQDETCLKAHLKLEVDLNSKLENTFLQELVELKMLAEEDKRRVSE--LTSKLT 1180

Query: 519 DTIQELQRELQMLQKSSMAEKEQTSNRKRVEELSLESEALRKLENSDKERQLQKTV 578
T+E Q +K + E + +K EEL+++L +K E +K + +

Sbjct: 1181 -TDEEFQSLKSSHEKSNKSLKEDKSLFKKLSLELAIQLDICCKTEALLEAKTN--ELIN 1237

Query: 579 EQDMKMNMDLRIKH-QHREQGSICKLEEDLQEATKLEDKREQLKSKSEHEKLMEGEL 637
K N +L RI H QHR K++E L T + + QL++ E + +

Sbjct: 1238 ISSSKTNAILSRIHQHRTT----KVKEALLIKTCTVSELEAQLRQLTEEQNTLNISF 1292

Query: 638 EALRQEFKKKD---KTLKENSRLKEEENENL-----AELQCCSTQLESSL---- 680
+ + ++K+ K++K + L E E L+ +E + C TQL+ L

Sbjct: 1293 QQATHQLEEKENQIKSMKADIESLVTEKEALQKEGNGQQAASEKESCITQLKKELSENI 1352

Query: 681 NKYNTSQVQIDLNKEIALQKESLMSLQALQDKALQ-KEKHYLQTTITKEAYDALSRKSA 739
N ++ +++ EI+ + L L QL ++ EK +++ K+ YD +

Sbjct: 1353 NAVTIMKEELKEKKVEISSKQLTDLNVQLQNSISLSEKAAISSLRKQ-YDEEKCELL 1411

Query: 740 ACQDDLTQALEKLN-HVTSETKSLQSSLTQTQEKKAQLEEEIAYEERMKKLWTELR-KL 797
DL+ ++ L+ S + + + + E K + + ++ +K+L +L K

Sbjct: 1412 DQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWKKQAQSRFTQHQTIVKLEQLQLELKS 1471

Query: 798 RGFHQESELEVHAFOKKLEEMSCQVLQWQKHQNDKMLAAKEEQLR-EFQEEAALKEN 856
+ +++ E +++ +L+++ + + + +D + KE L E + + A + E

Sbjct: 1472 KEAYEKDE-QINLLKELDQKNRFDCLKGEMEDDKSKMEKKESNLETLSQTARIME- 1529

Query: 857 LLEDDKEPCCLPQWSPKDTCLRYGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIA 916
LED + + T + N+ ++ N Q OK K +L +++ +

Sbjct: 1530 -LEDH-----ITQRTIEIESLNE-VLKNNYQ---QKDIENK---ELVQKLQHFQ 1570

Query: 917 KLSGKDKH----LHSVVMVHLQOENKKLKEIEKKMKAENTRLCTKA 959
+L EKD+ ++ L+ + +K E+K K + E+ L K+

Sbjct: 1571 ELGEEKDNRVKAEAEKILTLENQVYSMAKAELETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22
Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMSSHDKKQAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQT 106
+ E N + + Q EE E + S K ++ L + LQ+E +
Sbjct: 1281 LTEEQNTLWISFQOATHQLEEKENQIKSMKA----DIESLVTEKEALQKEGGNQQQAASE 1336

Query: 107 SYYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNENTGEKHLAQEQIAL 166
+ Q + L + + + L+ K K+ E+ +++ N + L++++ A
Sbjct: 1337 KESCITQLKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLRYDKYQSSLSNIELLECQVKMLQGELEGGIMGQEPENKGDHKSRIY 226
I+SL + Y ++ L ++ L +V L E + Q + S+ +
Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVMQKVSQDDLIQEL--RNK-LACSNALVLE--- 272
+ HQ T K L E ++K Q + L +EL +NK C + +
Sbjct: 1448 KAQSRTQHQONTVKELQIQLELKSKEAYEKDEQINLLKEELDQONKRFDCCLKGEMEDDKS 1507

Query: 273 -REKALIKLQADFASCTAT---HRYPSSSEECEDIKKILKHLEQKDSQCLHVEEYQN 327
EK L+ + S TA + + E + ++LK+ +QKD E++
Sbjct: 1508 KMEKESNLETLSQATARIMELEDHITQKTIEIESLNEVLKNYNQQRDI-----EHKE 1561

Query: 328 LVKDLRVELEAVSEQKRNIMKDMKLELDLHGLREETSABIERKKDII--TILQCLQEL 385
LV+ L+ + + E+K N +K+ + L L A +E K K++ L + +E
Sbjct: 1562 LVQKLQ-HFQELGEEKDNRVKEAEKILTLENQVYSMAELETKKKELEHVNLSVKSKEE 1620

Query: 386 QLETTETQKLTLLKDKFLQEKDEMLOLEKKLTQVQNSLLKKEKELEKQCMATELEMTV 445
+L+ E + L+ + + E+ ++ E+K+ ++ LL + +E +Q TE ++
Sbjct: 1621 ELKAEDR---LESES-AAKLAELKRKAEQKIAATKQLLSQMEKEEQYKKGTEHSLSE 1676

Query: 446 KEAKQDKSKEAECKALQAEVQKLNKSLSEAKQOERLAAQQAQCK-EAALAGCHLEDTQ 504
K + +E E L+ +++ ++S E R A AA + EEA GC + +
Sbjct: 1677 LNTKLQE-REREVHILEEKLSVESSQSETLIVPRSAKNVAAYTEQEEADSQGCYQKTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQRELQMLQKSSMAEKEQTSNRKRVLESLSEALRKL 564
K+ +L + + + LQR Q +KE +++ + R + +E ++L A K
Sbjct: 1736 EKIS---VLQRNLTEKEKLLQVRGO--EKEETVSSHFM--RCYQERLIKLEHAEAKQH 1788

Query: 565 NSDKERQLOKTVAEQDMKMDMLDRIKHQHQREG--SIKCK--LE---EDLQ-----E 611
LQ+ + E++ K + ++ +H +E G +I+ K LE +D+Q E
Sbjct: 1789 EDQSMIGHLQEELEENKKYSLIV--AQHVEKEGGKNNIQAKQNLNVFDDVQKTLQKE 1846

Query: 612 AT-KLEDKREQLKKSKEHEKLMEG-ELEALRQEPKKDKTLKENS---KLEENENL 665
T ++LE K ++L +K + E+ L + +K + + R +L EEN
Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVMEELTSKYELQALQMDGRNKPTELLEENTEE 1906

Query: 666 RAELOCCSTQLESSLN-KYNTSQOVIQDLNKEIALQKESLMSLQALQDKALQEKHYLQ 724
++ +L S++ ++N + + +E + ++ LQ L + L+KE H +
Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKIVRLQKDL-RMLRKE-HQQEL 1964

Query: 725 TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQOSSLTQOTQEKKAQLEEEIAYE 784
I K + YD R+ Q+ + LE L H ++ + ++ TQ +K+ +LE I +
Sbjct: 1965 EILKKEYDQ-EREKIKQEQ--EDLE-LKHNSTLKQLMREFNTLQAEQEQLEMTI---K 2017

Query: 785 ERMKKLNTLRLKRGFHOSELELVHAFDKKLEEMSCQVLOWQKQHQNDKMLAAKEEQLR 844
E + K +L HQE E + KK+ E + + K+++ ++L A+EE++
Sbjct: 2018 ETINKAQEVEAEELSHQE---ETNQLLKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEAALKENLLEDDKEPCCLPQWSVP-KDTCRLYRGNDQIMTNLEQWAKQKQVAN 903
++ E L + ++ L Q P D + ++ T L Q K ++ K
Sbjct: 2072 AKVRDLQTLQLEELQKQYQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTLISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDLHLSVMV-HL 932
L Q REQ++ + +L + ++++ V HL
Sbjct: 2124 LKEQEFREQIHNLDRLLKYEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20
Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQDIKN-LHDVC-KRQKTLQDNQCMEEAM-NSSHDKK 58
++ ++ E+ +L KLL + +K L + + +K Q N +E A NS+
Sbjct: 957 EKVQKAKEMQETLKKLLDQEAKKKELENTALELSQKEKQFNKMLEMAQANSAGISD 1016

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYYSLRQYQSIL 118
L + E + S + H R+L + + + ++L EELO + ++ +
Sbjct: 1017 AVSRLETNQKE-QIESLTVHRRRELNDV---ISIWKKLNQQAEELO-EIHEIQLQEK-- 1069

Query: 119 EKQTSDLV--LLHHHCKLKE-DEVILYEEEMGNHNENTGEKHLAQEQIALAGDKIASLE 175
E++ ++L +L C+ +E ++ I + +E G + T +L +Q + + +A E
Sbjct: 1070 EQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDITLNELOEQKQSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVKMLQGELEGGI--MGQEPENKGDHKSRIYTSPCMIQ 233
L + +K + L N L E LQ +L + + +E + K ++ T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQLVELKMLAEEDKRKVSELTSKLKTTDEEFQ 1186

Query: 234 E---HQETQKRLSEVWQKVSQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNKSLD---KSLEFKKLEELATQDICKKTEALLEANTELINISSKT 1243

Query: 286 SCTATH-RYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKR 344
+ + + + + + + + I + + + Q + E QN + + + E + K

Sbjct: 1244 NAILSRISHCQHRITTKVKEALLINTCTVSELAQLRQLEEQNTLNISFQQAATHQLEKE 1303

Query: 345 NIMKDMKLELD-LHGLREETSATIERKDKDITILQCRLEQLQLETTET-QKLTLLKDKF 402
N + K M K + + L + E + + + + + + + L + E + E + + TL K + +

Sbjct: 1304 NQIKSM-KADIESLVTEKALQKEGNOQQAASEKESCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQEKDEMLOLEKLLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKSKEAECKALQ 462
L + EK + L K + LT + N L + L + + + + L E K + . + + L

Sbjct: 1362 LKEKKVEISSLSKQLTDL-NVQLQNSISLSEKAAISSLRKQYDEEKCELLOQVQ--DLS 1418

Query: 463 AEVQKLKNSLEEAQOERLAAQAAQCKEEAALAGCHLEDTORLKLQGLLLDKQKA---- 518
+ V L A + Q + + + + K + A + + T + L Q L L + A

Sbjct: 1419 FKVDLTSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLEKSKAEYKED 1478

Query: 519 DTIQELQREQLMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKERQLOKQTV 578
+ I L + EL K + E + + + E + L + L + + + L + +

Sbjct: 1479 EQINLLKEELDQQNKRFDCLEKGEEDDKSKMEKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMDMLDRIKHQHREQGSIKCK-LEEDLQEAATKLEDKREQLKSKSEHEKLMEGEL 637
+ + + + + + + K + + Q I + K L + L Q + L E + K + + K + + E + E + +

Sbjct: 1536 QKTIEIESLNEVLKN-YNOQKDIEHKELVQKLQHFQELGEEKONRVKEAEKILTLENQV 1594

Query: 638 EALRQEFKKDKTLKENSRRLEENENLRAELQCCSTQLES-SLNKYNTSQQVQIDLNKE 696
+ + + E + K K L + + + + E L + A L + + LES S K + + + +

Sbjct: 1595 YSMKAELETKKKELEHVNLSVKSKEELKA-LE---DRLESESAAKL---AELKRKAQK 1647

Query: 697 IALQKESLMSLAQALDKALQKEXHYLQTTITKEAYDALSRKSAACQDDLTQALEKNHVT 756
IA K + L + S Q + + + KE + Y + T + L + K + + + + EKL V

Sbjct: 1648 IAAIKKQLLS---QME---EKEDQYKKG--ESHLSLNTKLQEREREVHILEEKLSVE 1699

Query: 757 S---ET---KSLQOSLTQTEKKAQLEEEII-AYEERMKKLNTLRLKRGHOFHOSELEV 808
S ET + S + T + + + A + + YEE + L L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEADSGCVQKTYEEKISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKLEEMSCQVLQWQKHQNDLKMALAAKEQLREFQEMAALKNLEDDKEPCCLP 868
+ + EE + + Q + Q L L E + E Q + L + E L E + K + +

Sbjct: 1753 QRVGQKEETVSSHFMRCQYQERLIKLEHAEAKQHEQDSMIGHLQEELEENKKYSLIV 1812

Query: 869 QMSVPKDTCLRLYRGNDQIMNTLEQ-WAKQKQVANER-LGNQLREQ-VNYIAKLSGEKDH 925
V K + + N Q NLE + OK EK L Q + EQ + + + +

Sbjct: 1813 AQHVEKEGGK---NNIAQAKNLENVDDVQKTQKEKELTCQILEQKIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQENKKLK 940
H V M L + + KL +

Sbjct: 1870 HRVEMEELTSKYELQ 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHQETQKRLSEVWQKVSQDDLIQE-LRNKLACSNALV-LEREKALIKL-QADFASCTA 289
+ E + TO + + + V + L + + L S + + L R + L + D S TA

Sbjct: 53 RESGDTQSFQAKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNMKD 349
+ P E ED+ L + + + Q L + + R + + + + +

Sbjct: 113 SFDPPSDMDSEADLVGNSDSLNEQLIQLRL--RMERSLSSYRGKYSSELVTAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSATIERKDKDIT-ILQCRLEQLQLETTETQKLTLLKDKKFLQEKDE 408
KL+ G+ + + + DK + I + R + ELQ + + L + D L + EKD +

Sbjct: 171 KKKLQ----GILSQS-----QDKSLRRIAEELR-EELQMDQQAQKHLQEEDFASLEEKDQ 219

Query: 409 MLQLEKLLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKSKEAECKALQAE---V 465
+ L + + + + L + + + + + LE + + + + E + + + + V

Sbjct: 220 YISVLQTVSLLKQRLRNGPMNVOLKPLP-QLEPQAEVFTKEENPESDGPVVEDGTSV 278

Query: 466 QKLKNSLEEAQOERLAAQAAQCKEEAALAGCHLEDTORLKLQGLLLDKQKADTI 521
+ L + + K + QE L + + Q KE+ L E Q + L + L L + K K +

Sbjct: 279 KTLQTLQORVQRQENLLKRCKETIQSHKEQCTLLTSEKALQEQLOERLQLEKIKDLHM 338

Query: 522 QELQREQLMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKERQLOKQTVAEQD 581
E + + L + + + E + + + E + + E L E + R K + Q

Sbjct: 339 AEKTKLITQLRDAKNLIEQLEQDKGMVIAETKQMHETLEMKEEIIAQLRSRIKQMTTQG 398

Query: 582 MKNMMDLRIKHQHREOGSICKLEEDLQEAT-KLEDKREQLK---KSKEHEKL-MEGE 636
 ++ ++ ++ E+ + +EA KL + EQ+K K+ E E++ ++ E
 Sbjct: 399 EELREQEKESERAAFELEKALSTAQKTEEARRLKAEMDEQIKTIKTSEERISLQOE 458

Query: 637 LEALRQEFKK-KDKTLKENSRLKEEENLRAELQCCSTQLESSLNKYNTSQVQIDLNK 695
 L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K
 Sbjct: 459 LSRVKQEVVDMKSSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLMSLQAQLDKALQKEKHLYLQTTITKEAYDALSRKSAACQDDLTQALEKLN-H 754
 +AL+K L+ +K Q+ + +K+A S DL Q E
 Sbjct: 513 -VALEKSQSEYLIKISQEKEQESLAELELQKKAILTESENKLR---DLQQAETYRTR 568

Query: 755 VTSETKSLQOQLTQOEKKAQLEEEIAYEERMKKLNTLRLKRGFHOESELEV--HAFD 812
 + SL++SL OE K Q ++ + E K N E+ + H+ +ELE H D
 Sbjct: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHKEITVMVEKHK-TELESKXHQD 624

Query: 813 KKEEMSCQVLQWQKHQNDLKMALAAKEEQLRE-----FOEMAALKENLLED-DK 862
 E QVL+ +Q+Q +++ L K EQ +E FO + + E LE D
 Sbjct: 625 ALWTE-KLQVLK--QQYQTEMEKLRKCEQEKETLLKDKEIIFQAHIEEMNEKTLEKLOV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANERKLGQLEQVNYIAKLSGEK 922
 + L S+ + + + + L Q ++L ++ EQ N+ +
 Sbjct: 682 KQTELE--SLSSELSEVLKARHKLEELSVLKDQTKMKQLEAKMDEQKNHHQQQVDSI 739

Query: 923 DHLHSMVMHLQOENKKLKEIEEKKM 948
 H V + Q+ K LK +I + ++
 Sbjct: 740 IKEHEVSI--ORTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09
 Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQEEKLLTQVQNSLLKKEKELEKQCCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467
 M ++L++K+++ Q L + + +T M + + + ++ E + + Q
 Sbjct: 1 MFKKLKQKISSEQQQLQALAPAQASSNSSTPTRMSRTSSFTQLEDGTPNRESGDTQS 60

Query: 468 LKNSLE-EAQQERLAAQQAQCKEEAALAGCHLEDTRKQLKGLLLDKQKA--DTIQEL 524
 L+ E L + + + + +R+ L LD A D ++
 Sbjct: 61 FAQLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTASFDPPSDM 120

Query: 525 QRELQMLQKESMAEKEQTSNRKRVEELS-----ELSEALRKLNSDKERQLOKTVAE 579
 E + L S KEQ R R E SL + SE + + +EK++LQ +++
 Sbjct: 121 DSEAEOLVGNSSSLNKEQLIQLRLRMERSLSSYRGYSELVTAYQMLQREKKKLGILSQ 180

Query: 580 -QDMKNMMDLRIKHQHREOGSICKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632
 QD + + + + +Q+ K EE L+E + +L+ + LK+ + +
 Sbjct: 181 SQDKSLRRIAELELQMDQAKKHLQEEFDASLEEKDQYISVLQTVSLLKQRLRNGPM 240

Query: 633 MEGEALRQ-EFKKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQ 688
 L+ L Q E + + T +EN E E+ L+ +++ N ++
 Sbjct: 241 NVDVLKPLPQLEPAEVFTKEENPESDGEFVVEDGTSVKTLETQQRVQRQENLLKRCKE 300

Query: 689 VIQDLNKEIALQKESLMSLQAQLDKALQKEKHLYLQTTITKEAYDALSRKSAACQDDLTQA 748
 IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +
 Sbjct: 301 TIQSHKEQCTLLTSEKALQEQLDERLQ-ELEKIKDLHMAEKTKLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQOQLTQOEKKAQLEEEIAYEERMKKLNTLRLKRGFHOESELE 807
 LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E
 Sbjct: 358 LEQDKGMVIAETK---RQMHTLENK---EEETIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKHQNDLKMALAAKEEQLREFQ-----EEMAALKENLLEDDKE 863
 AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E
 Sbjct: 410 RAAF---EELKALSTAQKTEEARRLKAEMDEQIKTIKTSEERISLQQLSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKQVANERKLGQLEQVNYIAK 917
 + S + +L + +++ + EQ K+ + + Q++ Q Y+ K
 Sbjct: 466 VVDVMKSSSEEQIAKLQKLHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDLHSMVMH-LQOENKKLKEIEEK---KMKAEENTRLCTKALGPSRTESTOREK 972
 +S EK+ S+ + L+ + K+ E E K + +AE R L S +S Q K
 Sbjct: 525 ISQEKEQESLAELELQKKAILTESENKLRDLQQAETYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3_lg13, frame 1

Report for DKFZphtes3_lg13.1

{LENGTH} 1007
 {MW} 117480.77
 {PI} 5.90

[HOMOL] TREMBL:AF092090_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04
[EC] 3.6.1.32 Myosin ATPase 1e-16
[PIRKW] nucleus 3e-10
[PIRKW] phosphotransferase 6e-09
[PIRKW] duplication 2e-06
[PIRKW] citrulline 2e-12
[PIRKW] tandem repeat 1e-16
[PIRKW] endocytosis 2e-13
[PIRKW] heart 8e-13
[PIRKW] transmembrane protein 1e-13
[PIRKW] serine/threonine-specific protein kinase 6e-09
[PIRKW] zinc finger 2e-13
[PIRKW] metal binding 2e-13
[PIRKW] DNA binding 4e-12
[PIRKW] muscle contraction 1e-16
[PIRKW] acetylated amino end 1e-11
[PIRKW] actin binding 1e-16
[PIRKW] mitosis 5e-15
[PIRKW] microtubule binding 5e-15
[PIRKW] ATP 1e-16
[PIRKW] thick filament 1e-16
[PIRKW] phosphoprotein 4e-16
[PIRKW] skeletal muscle 2e-14
[PIRKW] calcium binding 2e-12
[PIRKW] alternative splicing 1e-16
[PIRKW] coiled coil 1e-16
[PIRKW] P-loop 1e-16
[PIRKW] heptad repeat 3e-10
[PIRKW] methylated amino acid 1e-16
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 2e-13
[PIRKW] cardiac muscle 8e-13
[PIRKW] hydrolase 1e-16
[PIRKW] microtubule 3e-10
[PIRKW] muscle 8e-13
[PIRKW] EF hand 2e-12
[PIRKW] cytoskeleton 2e-15
[PIRKW] hair 2e-12
[PIRKW] calmodulin binding 2e-13
[PIRKW] Golgi apparatus 3e-10
[SUPFAM] myosin heavy chain 1e-16
[SUPFAM] conserved hypothetical P115 protein 1e-07
[SUPFAM] centromere protein E 5e-15
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09
[SUPFAM] calmodulin repeat homology 2e-12
[SUPFAM] myosin motor domain homology 1e-16
[SUPFAM] alpha-actinin actin-binding domain homology 2e-07
[SUPFAM] plectin 2e-07
[SUPFAM] trichohyalin 2e-12
[SUPFAM] pleckstrin repeat homology 8e-08
[SUPFAM] ribosomal protein S10 homology 2e-07
[SUPFAM] giantin 3e-13
[SUPFAM] protein kinase homology 6e-09
[SUPFAM] protein kinase C zinc-binding repeat homology 8e-08
[SUPFAM] kinesin motor domain homology 5e-15
[SUPFAM] human early endosome antigen 1 2e-13
[SUPFAM] M5 protein 1e-07
[PROSITE] LEUCINE_ZIPPER 7
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 20

[illegible]


```
COILS .....
SEQ  NEKLGNQLREQVNYIAKLSGEKDHLSVMVHLQENKKLKEIEKKMKAENTRLCTKAL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ  GPSRTESTQREKVCGLGWKGLPQDMGQRMDLTKYIGMPHCPGSSSYC
SEG  .....
PRD  cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

Prosites for DKFZphtes3_lg13.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00001	684->688	ASN_GLYCOSYLATION	PDOC00001
PS00004	240->244	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	415->419	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	110->113	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	396->399	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	503->506	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	566->569	PKC_PHOSPHO_SITE	PDOC00005
PS00005	600->603	PKC_PHOSPHO_SITE	PDOC00005
PS00005	650->653	PKC_PHOSPHO_SITE	PDOC00005
PS00005	655->658	PKC_PHOSPHO_SITE	PDOC00005
PS00005	735->738	PKC_PHOSPHO_SITE	PDOC00005
PS00005	876->879	PKC_PHOSPHO_SITE	PDOC00005
PS00005	968->971	PKC_PHOSPHO_SITE	PDOC00005
PS00006	39->43	CK2_PHOSPHO_SITE	PDOC00006
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	444->448	CK2_PHOSPHO_SITE	PDOC00006
PS00006	471->475	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	566->570	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->580	CK2_PHOSPHO_SITE	PDOC00006
PS00006	650->654	CK2_PHOSPHO_SITE	PDOC00006
PS00006	674->678	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	888->892	CK2_PHOSPHO_SITE	PDOC00006
PS00006	963->967	CK2_PHOSPHO_SITE	PDOC00006
PS00006	968->972	CK2_PHOSPHO_SITE	PDOC00006
PS00007	135->143	TYR_PHOSPHO_SITE	PDOC00007
PS00008	207->213	MYRISTYL	PDOC00008
PS00008	599->605	MYRISTYL	PDOC00008
PS00029	83->105	LEUCINE_ZIPPER	PDOC00029
PS00029	90->112	LEUCINE_ZIPPER	PDOC00029
PS00029	97->119	LEUCINE_ZIPPER	PDOC00029
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029
PS00029	403->425	LEUCINE_ZIPPER	PDOC00029
PS00029	410->432	LEUCINE_ZIPPER	PDOC00029
PS00029	918->940	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_lg13.1)

DKFZphtes3_1k11

group: cell structure and motility

DKF2phtes3_lkl1 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, compete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```
1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGGCGG GGATCGTCTC
101 CGGCACTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGCG
151 GAGCAGCAGC GGGTCCATGA ACGTCAACCTT CTCCACAAG GCCTCCACCC
201 CGGACTGTGT GCTGGCCACG CTCACACGCG TTCCCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTGGGC GGGCGACCGT GCCTTCCCTT GTACCCGTGC
301 CGTGTGGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTT AGCCATGGCC
351 TTCCGGGAGG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCTCAC GCATCGCCAT
451 CAACGAGGAG AACGCTGAGT CACTGTCTGA GGCAGGCGAC ATGCTGCAGT
501 TCCACGATGT GCGGGATGCT GCGCGCGAGT TCCTGGAGAA GAACCTTTTC
551 CCCTCCAACCT GCCTGGGCGT GATGCTGCTC TCGGAGCGCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCCT GGCGCATGTG CTTGTGTCAC TTTCAGACGG
651 TGAGGCGAGG CGAGGACTTC AACAGCTGTG CCAAGCACAC ACTGCTGGAC
701 CTCATCTCCA GTGATGAGCT GGAGACCGAG GACGAGCGGG TGCTCTCGA
751 GGCCATCCTC CAGTGGGTGA AGCAGCACTT GGAGCCACGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGACG GTGCGTCTGG CTTGTCTGCC GTCCGACTGC
851 CTGAGGAGGG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACGAGG ATCCTGCAGA
951 ATGATGGCGT GGTCAACAGC CCCTGTGCCC GGCCACGCAA GCGGGGCCAC
1001 ACGCTACTCA TCCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACAC AAGGCCAAGG AGATCATCCC CAAGGCCGAG CTGCCACGCC
1101 CCCGGAAGGA GTTCAGCGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGGTCTCC AAGGATGCTT GGTGTACGGA
1201 CACCGTACAT GAGGAATGGT CCAAGGCGCG GCCATGCTGT ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTTCGCCGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGGCGGA TGGCGTCAGC AATGCCGCAG TGGTGAATGC CAAGCTGAAG
1451 CTCTTTGTTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAAGTCTAT GACCCCTCGG AGAACAGGTT GACGATCAAG GCCGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCGCTGCCC TCCTGGGCAG CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTACAGCCG GCCTCGGCCCT ACCGCTTTGA
1651 CTGTGAGACC AACCAGTGGG CGCGGATTGG GGACATGACT GCCAAGCGCA
1701 TGTCTTGCCA TGCCCTGGCT TCCGGCAACA AGCTCTATGT GGTGCGGGC
1751 TACTTTGGGA CCCAGAGGTG TAAGACTCTG GACTGTATGT ACCCCACTTC
1801 AGATACATGG AACTGCATCA CACAGTGGCC CTACTCACTT ATCCCCACGG
1851 CTTTGTGAG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCCAGCCAGA CCGCGCCCTT CAGTGTCA CACTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGGTGGAAGG
2001 GGGCCTGCCA GCTCTGGGGA GCAGCAGCCT TGGGCTGTTT TGAGCTTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTCATCCG TGCCCTGGGG GGCCTTTTCA
2101 GCTTTGAGAG GGTTTGTGGG AAGACATACC TCCCAGAGGG GCATGGAGTG
2151 CCACCAAGAG TGACCTTGGC GTCCGGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACTGT TTTGGCAGGT CCTGGACTGG GGGCGGGCAG GCAGGGGCGAG
2251 GGAGGCGCCC CGGGTGGGCT TTTGGGCTGC GGCATCTAT CACATCCTTT
2301 CCCTCTGGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGTG
2351 CTGGGCTGGG GAAACTAGGT TCCAGGGGGT TGAGACGAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAACTGCA GGGGGGCTC AGTGACATCC
2451 ATGAGGCCCTT ATTAGCAAAG GACACCCAGA CCTCCAAGGT TTGTGGGCCC
2501 CTTCACAAA GCTGTAAGTC CCAGCCACCC TACTCAGGGC CTTGCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCCTCCCT
2651 CCCCCTCTCA GAGCCACCC TGAGAGGCGA CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GCTTCTCCCC AGGAGACCAC GCTGCCCTCC
```

```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCCCT CACCATCCTC TTGCGTTTTG GCCCCCATTT GCTCCCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCAATTGC
2951 TGTCACCTCC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTGGCG CCTGTCACTG ACCATCCTAA TGCCTCTCGC
3051 TCACTCCCAA GCCACCATTT GAGAGGGAGG GGTGTTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC CGGAGGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCCTC TTGGGGGTGG GACCTTTGGG CAGGTTTGGC CCACTGACGC
3301 GCCCGCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGTTGCCTG GACCCGTGGG
3401 GTTGTTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTTGCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

98350113:
Cloning of human ENC-1 and evaluation of its expression
and regulation in nervous system tumors.

97252647:
ENC-1: a novel mammalian kelch-related gene specifically expressed in
the nervous system
encodes an actin-binding protein.

98234394:
NRP/B, a novel nuclear matrix protein, associates with
p110(RB) and is involved in neuronal differentiati

Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589
Category: strong similarity to known protein
Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLTNR KHCNFTDVTIL
51 WAGDRAFFCH RAVLAASSRY FEAMFSLGLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAVSSRI AINEENAESL LEAGDMLQFH DVRDAAEFL EKNLFPSNCL
151 GMLLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK DTLDLIISD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEA
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGQTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVYVGGGGS
351 ENGVSKDVMV YDVTVEEWSK AAPMLIARFG HGSAELENCL YVVGGHTSLA
401 GVFPASPSVS LKQVKEYDPG ANKMMMVAPL RDGVSNAAVV SAKLKLFPVG
451 GTSIHRDMVS KVQCYDPSFN RWTIKAECPQ PWRYTAAAVL GSQIFIMGGD
501 TEFTAASAYR FDCETNQWTR IGDHTAKRMS CHALASGNKL YVVGGYFGTG
551 RCKTLDCYDP TSDTWNCITT VPYSLIPTAF VSTWKHLPA

```

BLASTP hits

Entry MMU65079_1 from database TREMBL:
gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds.
Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611_1 from database TREMBLNEW:
gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens
nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.
Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314_1 from database TREMBL:
gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,
complete cds.
Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN). >TREMBL:DMRCPA_1 product: "ring canal protein"; Drosophila melanogaster ring canal protein and ORF2 mRNA, complete cds.
Score = 672, P = 3.9e-66, identities = 168/536, positives = 257/536

Alert BLASTP hits for DKFZphtes3_1kl1, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_1kl1, frame 2

Report for DKFZphtes3_1kl1.2

[LENGTH] 589
[MW] 65923.45
[pI] 6.10
[HOMOL] TREMBL:MMU65079_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds. 0.0
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
2e-09
[BLOCKS] BL01016D Glycoprotease family proteins
[PIRKW] zinc finger 1e-08
[PIRKW] DNA binding 1e-08
[PIRKW] transcription factor 1e-08
[SUPFAM] POZ domain homology 3e-68
[SUPFAM] vaccinia virus 59K HindIII-C protein 1e-15
[SUPFAM] A55R protein 5e-29
[SUPFAM] hypothetical protein YHR158c 4e-08
[SUPFAM] A55R protein middle region homology 5e-29
[SUPFAM] myxoma virus M9-R protein 1e-14
[SUPFAM] A55R protein carboxyl-terminal homology 5e-29
[KW] Alpha_Beta

SEQ MSVSVHETKRSRSTGSMNVTLFHKASHPDCVLAHLNLTLRKHCMTDVTLMAGDRAFPCH
PRD ccc

SEQ RAVLAASSRYFEAMFSGHGLRESRDDTVNFQDNLHPEVLELLLOFAYSSRIANEENAESE
PRD hcc

SEQ LEAGOMLQFHVDRAAAEFLEKNLFPSNCLGMMLLSDAHQCRRLYEFSWRMCLVHFETVR
PRD hhh

SEQ QSEDFNSLSKDTLLDLISSDELETEDERVVFEAILQWVKHDLPRKVHLPPELLRSVRLAL
PRD hhh

SEQ LPSDCLQEAVSSEALLMADERTKLIMDEALRCKTRILQNDGVVTPCARPRKAGHTLLIL
PRD ccchhh

SEQ GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVWV
PRD ccc

SEQ YDTVHEEWSKAAPMLIARFGHGSAELENCLYVVGHTSLAGVFPASPSVSLKQVEKYDPG
PRD ccc

SEQ ANKWMHVAPLRDGVSNAAVVSAKLKLVFGGTSIHRDMVSKVQCYDPSENRTWIKAECPQ
PRD ccc

SEQ PWRYTAAAVLGSQIFIMGGDTEPTAASAYRFDCEWTRIGDMTAKRMSCHALASGNKL
PRD ccc

SEQ YVVGGYFGTQRCKTLDCYDPTSDTWNCITTVPSLIPTAFVSTWKHLPA
PRD ecc

(No Prosite data available for DKFZphtes3_1kl1.2)

(No Pfam data available for DKFZphtes3_1kl1.2)

DKF2phtes3_in3

group: signal transduction

DKF2phtes3_in3 encodes a novel 1196 amino acid protein with similarity to *S. pombe* Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```
1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAAACTTG TCAGGTCTGA AGAAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACACAAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCATATT AAAGAACTA
251 CAAGTGATGA TGAAGTGCT GCTAACCTA ACAACCTGAA GAAGAGCAGC
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAATAAAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGAAAT TGATGCAAGC ATATCAATGC
601 CATGTAACCTG AAGAATGCG AAAGGAGATT AAGAGGAAA TAAGAAAGAA
651 ACTGAAAGAA CAGTTGACTT ACTTCCCTC AGATCTTTA TTCCATGATG
701 ACAAACTAAG CAGTGAATAA AGGAATAAGA AAAAGGAAGT TCCAGTCTTC
751 TCTAAGCTG AAACAAGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
801 TGAACAAAAG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGTCA ATGGAACAAA GCACAGAAGA CAGCATGCAG
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAGAAGA CTAAGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTCAT GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTACCG AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGGTGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTCTATCTA CTATGAAAAA
1201 GAGAAATGGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTGTGCT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAATT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAATG CAAACATCAA CTCAAAACTT
1501 CGCTTGACGC TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAAGTG
1551 TGTTGAGGCA TTTGAATGGT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACTGTA AGAGGACTGA AAGTTCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGTCTCTCAG GAGGAAAAG GTAAACCAAT
1701 GCATTGTGAA CGTCACCATG AGTCAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGG AACGACTCCC TGGCAGGCT
1801 TGCCGATATCC CAACAAACA CCTCTTCTCA CTAATGCAG GAGACGAGG
1851 ATGTTTGTGT CTGATTCTCT CCCACAATGG AAGAAATATA GCACAGCTT
1901 GTGCCAGCCG GATGGATAT CCAATTATTT TATATGAAAT TCCTCTGGA
1951 CGTTTCATGA GAGAATTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGGTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACCTT CAGAGTTTAA
2101 CCTCATCCTT CTTTGTGTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTGATACT GAAGGTATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAAATTA
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTGGAGA TTCATCCCAA
2451 TGGAAACGCT TTGTTAATCC ATACCAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTTATCGG
2551 GAGAAGATTC ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTCGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGAA CCCAGAAACA GGAGAACAAG
```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTGG AGACATTCTT
2701 TATCATCCAT TTGAAAATAT GGTTCGATTC TGTGCAATTC GGCAAAATGA
2751 GCCAATTCCT CTGTATATTT ACGATTTCGA TGTGCCCAG CAGGAGGCTG
2801 AATGTTCAA ACGCTACAAT GGAACATTTC CATTACCTGG AATACACCAA
2851 AGTCAAGATG CCCTATGTAC CTCTCCAAA CTACCCCATC AAGGCTCTTT
2901 TCAGATTGAT GAATTTGTCC ACACGAAAG TTCTCAACG AAGATGCGAG
2951 TAGTAAACA GAGGCTTGAA ACTGTACAG AGGTGATACG TTCTGTGGCT
3001 GCAAAAGTCA ACAAATCT CTCAATTACT TCACCACAG CAGTTTCCTC
3051 ACAACAGTCT AAGTTAAAGC AGTCAAACAT GCTGACCGCT CAAGAGATTC
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAGAGAAAG
3151 CCTTGTAAAC ATCAGGTAGA TACAGACCA ACGGTAGTGG CTCTTTATGA
3201 CTACACAGCG AATCGATCAG ATGAACAAAC CATCCATCGG GGAGACATTA
3251 TCCGAGTGT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAACACT
3351 GTATCAAGAA CTGCCTCCTG AGATAAAGGA GCGATCCCTT CCTTTAAGCC
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA
3451 ATCAATAAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC
3501 ACATTCTGAA ATGAGAAAAG AACAGAGCCA TGAGGACCAA GGACACATAA
3551 TGGATACACG GATGAGGAAG AACAGCAAG CAGGCGAGAA AGTCACTCTA
3601 ATAGAGTAAA GAATTGAAGA AAGTTAAGA GCTGCCGAAA TGCACAGAGG
3651 TGAATAAGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCAG
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTGTCTCTT ATGATGACT
3751 CTAGAAAAAT CAGAAACAG TGTGGGTGG AAAAATCAAC GTGCCCTTTG
3801 AGTTCAGTGG TTATAACCA TTGTGACTAT TGTGGTCAA GTTATGGTA
3851 CTTATATTGT TACTAATTGC ATCATAATTA CATTACCAGT GTTGAANAAC
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAATATGTA ATAAAGGTG
3951 TTTGCGTTAT TAGGATGCT GTTAAGTAAT CATTAAATAT TATTATATTG
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT
4051 TGAATTCAC TTTATTAAAG AGATAAGCAG CTGACTGGGC ACGGTGCCCTC
4101 ATGCGCTGAA TCCTAGCACC TTGGGAGGCT GAGGCGAGGT GATCACCTAA
4151 GGTGAGGAGT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT
4201 CTACTAAAAA TACAANAATC AGCCGGGTCT CATGGCAGGC ACCTGTATATC
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTGCGAG
4301 TGAGCCAAAG TCACGCCATT GCACTCCAGC CTGGGGGACA GAGCAAGACT
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTAGAATA TGGCGCATTC
4401 AAAACAGTCT CAGTAACAAA GACATTAATA GAAAACAATT TACTTTCTAA
4451 TTAAAAATTT GTGTTCTTTA AGATCAAAAT ATATAGGTAA CTTATAGAC
4501 CTAAAAATAA AGTGATTTT GGCTGGACTG GCAACAATGT TCCCAATGTC
4551 TTTACTTTTT AAAAAAGGCT TTTCATATTT AAGCACATAC CTATTTTGTA
4601 GACTTACATT GTTTAATATT TATTTTAAAT TTAATATTTT TACATATTA
4651 TATTGCATTA TTTATTTTCT CTAAGTTCCA GAATAATAGT GTCATTTATTA
4701 TAGACTATAT GTTTTGAAGT TTGATATTAT AATGGGATAT TCATTTTGTG
4751 TCTTTCTCTT GACTCCTTTC TCAAGTGTGT TAAAGGTCT GCTGATAAAA
4801 TATTTAACCC CAAGAAAGTG AAAACTAATA TAAATTAGA AAGACCTATC
4851 CAAATTAGAC AGTCAATTC ATTAATAATA GAAGTGAGAA AAACAATGTT
4901 GGGCATTGAG GTGTAAATTT TGCCAGATG TATACCCAGT GTGAAATATC
4951 TTCTAATAAA AATATATTTT GCTCTTATCC CTGCACATGT AGAGGCATAA
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAAA AAGGCATTTT
5051 TGAAGAGTGT GAGTGCCACT GATAACTGGT GAAGCCTACA GCCATCCGCC
5101 CAAAAGTCTG TTCTGATGCG ACTGAGTTTT CATTGTTCTG GATGTATAAG
5151 TCTGTGTGTC AGGTACAGCT GGGCCAGGCC AGCTTGAGTC ACTCTGTATC
5201 AAGCTTGTGT TTTCTGTCT TGTGAATGCA CTTGATAATT TAAAAATAAA
5251 AATATCTGTT TCTCTGCAA AAAAAA

BLAST Results

Entry HS3281 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 32B1
Score = 4445, E = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:
Human exon-trapped sequence from 6q24.
Score = 965, E = 4.0e-35, identities = 193/193

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196
Category: similarity to known protein

```

1 MPTAESEAKV KTKVRFKELL KTHSOLMREK KKLKKLVRS EENISPDITR
51 SNLHYMKETT SDDPDTIRSN LPHIKETTSO DVSAANTNNL KKSTRVTKNK
101 LRNTOLATEN PNGDASVEED KQCKPNKKVI KTVPQLTTQD LKPETPENKV
151 DSTHOKHTHK PQPGVDHQS EKANEGREET DLEDEELMQ AYQCHVTEEM
201 AKEIKRKIRK KLKEQLTYFP SDTLFHDDKL SSEKRAKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHQDD EISSMEQSTE DSMQDQTKPK
301 PKTKKKKTKA VADNNEDVDG DGVHEITSRD SPVYFKCLLD DDLVLGVYIH
351 RTDRLSKDFM ISHPMKIHV VDEHTGQYVK KDSGRPVSS YYEKENVDYI
401 LPIMTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDES PKVILFFEIL
451 DFLSVDEIKN NSEVQNQECG FRKIWAFLK LLGANGNANI NSKLRLLQLYY
501 PPTKPRSPLS VVEAFVWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCERHHESS SVDTEPGLEE SKEVIKWRRL PGQACRIPNK
601 HFLFSNAGER GCFCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNNTNT FRVLPHPSFV
701 YTAKFHFAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CFDTGEGHMY SGDCGVIVV WNTYVKINDL EHSVHWTIN KEIKETEFKG
801 IPISYLEIHP NGKRLLIHTK DSTLRIMDLR ILVAREFVGA ANYREKIHST
851 LTPCGTFLFA GSEDIYVYV NPETGEQVAM YSDFEFKSPI RDISYHPFEN
901 NVAFCAFGQW EPIILYIYDF HVAQQAEMF KRYNGTFPLP GIHQSQDALC
951 TCPKLPHQGS FOIDEFVHTE SSSTKMQLVK QLETVTEVIE RSCAAKVNKN
1001 LSFTSPPAVS SQSKLQKSN MLTAQEILHQ FGFTQTGIIS IERKPCNHQV
1051 DTAPTVALYV DYTANRSDEL TIHRGDIIRV FFKDNEQWYV GSIGKGQEGY
1101 PFANHVASET LYQELPPEIK ERSPLSPEE KTKIEKSPAP QKQSKNNKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMRKNKQAGR KVTLIE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_1n3, frame 1

TREMBL:U92792_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CER03E1_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:Y2LL CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds.
Length = 321

HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18
Identities = 59/225 (26%), Positives = 111/225 (49%)

```

Query: 647 MRELCGHLNIIYDLSWSKDDHYILTSSSDGTARIWKNEINNNTNTFRVLPHPSFVYTAKFH 706
      + E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +P+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD--LLSASMDKTVRLW--QVGRDSCLKVFSTHNYVTCVOFN 119

Query: 707 PAVRELVVVTGCDYSMIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHMYSGDCTG 766
      +TGC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITGCDLVRIMDVRR-----CLVVDWANSKEIVTAVCYRPGKGAVAGTITG 174

Query: 767 VIVVMNTYVVKINDLEHSVHWTINKEIKETEFKGIPIISYLEIHPNGKRLLIHTKSTLRI 826
      ++ +LE V ++N K ++ Y P K+L++ + D+ +RI

```

Sbjct: 175 NCRYYDASENRLEESQV---SLNGRKKSLHKRIVGFQYCPSPDP--KKLMVTSGDAQVRI 229
 Query: 827 MDLRILVARKEVGAANYREKIHSTLTPCGTFLFAGSEGDIVYVWN 871
 +D +++ + G + + + + TP G + + S+D +Y+WN
 Sbjct: 230 LDGAHVSN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIMWN 272

Pedant information for DKFZphtes3_ln3, frame 1

Report for DKFZphtes3_ln3.1

[LENGTH] 1196
 [MW] 137114.70
 [pI] 6.79
 [HOMOL] SWISSPROT:YKY4_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN
 C14B1.4 IN CHROMOSOME III. 8e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
 TAF90 - TFIID subunit] 4e-10
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit]
 4e-10
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c]
 9e-08
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YDL145c] 9e-08
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YMR116c] 4e-06
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w]
 4e-05
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05
 [BLOCKS] BL00024H
 [SCOP] dltbgs_2.46.3.1.1 betal-subunit of the signal-transducing 3e-91
 [SCOP] dlqfc_2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14
 [SCOP] dlfmk_2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15
 [SCOP] dlad5b1_2.21.2.1.7 (1-63) Hemopoietic cell kinase Hck [human (Hom 3e-15
 [SCOP] dlckal_2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13
 [SCOP] dlqwea_2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15
 [SCOP] dlshg_2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13
 [SCOP] dlprmc_2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15
 [SCOP] dlhsq_2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13
 [SCOP] dlabo_2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13
 [SCOP] dlafna_2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15
 [SCOP] dlsema_2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13
 [SCOP] dlgbqa_2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16
 [SCOP] dlckaa_2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15
 [EC] 3.1.4.3 Phospholipase C 2e-07
 [EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07
 [EC] 3.6.1.32 Myosin ATPase 7e-07
 [EC] 2.7.1.112 Protein-tyrosine kinase 8e-06
 [PIRKW] nucleus 2e-08
 [PIRKW] phosphotransferase 8e-06
 [PIRKW] plasma 4e-07
 [PIRKW] duplication 4e-07
 [PIRKW] phosphoric diester hydrolase 2e-07
 [PIRKW] tandem repeat 7e-07
 [PIRKW] hormone 4e-07
 [PIRKW] transmembrane protein 2e-06
 [PIRKW] stomach 4e-07
 [PIRKW] actin binding 7e-07
 [PIRKW] ATP 7e-07
 [PIRKW] phosphoprotein 7e-07
 [PIRKW] signal transduction 7e-09
 [PIRKW] heterotrimer 7e-09
 [PIRKW] P-loop 7e-07
 [PIRKW] hydrolase 7e-07
 [PIRKW] transcription regulation 5e-06
 [PIRKW] GTP binding 7e-09

[SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07
 [SUPFAM] SH3 homology 2e-07
 [SUPFAM] SH2 homology 2e-07
 [SUPFAM] protozoan myosin heavy chain IB 7e-07
 [SUPFAM] myosin motor domain homology 7e-07
 [SUPFAM] pleckstrin repeat homology 2e-07
 [SUPFAM] protein-tyrosine kinase src 8e-06
 [SUPFAM] WD repeat homology 3e-12
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07
 [SUPFAM] protein kinase homology 8e-06
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-09
 [SUPFAM] yeast coatomer complex alpha chain 4e-07
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 19
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] Src homology domain 3
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW COMPLEXITY 5.77 %
 [KW] COILED_COIL 2.42 %

SEQ MPTAESEAKVTKVRFEKLLKTHSDLMREKKKKKKLVRSEENISPDITIRSNLHYMKETT
 SEGXXXXXXXXX.....
 COILSCCCCCCCCCCCCCCCCCCCCCCCC.....
 IgotB
 SEQ SDDPDITIRSNLPHIKETTSDDVSAANTNLLKKSTRVTKNKLRLNTQLATENPNGDASVEED
 SEG
 COILS
 IgotB
 SEQ KQKPKKKKVIKTVPLTTQDLKPETPENKVDSTHQKTHTKPQGVHDQKSEKANEGREET
 SEGXXX
 COILS
 IgotB
 SEQ DLEDEELMQAYQCHVTEEMAKEIKRKIRKKLKEQLTYFPSDTLPHDDKLSSEKRRKKKKE
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 COILS
 IgotB
 SEQ VPFVSKAETSLTISGDTVEGEQKKESSVRVSSDSHQDDEISSMEQSTEDSMQDDTKPK
 SEGXXXXXXXXXX.....XXXX
 COILS
 IgotB
 SEQ PKTKKKTKAVADNEDVDGCVHEITSRDSVPYPKCLDDDLVLGVYIHRTRLKSDFM
 SEGXXXXXXXXXX.....
 COILS
 IgotB
 SEQ ISHPNVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTPQYDFKQLKSRLPEW
 SEG
 COILS
 IgotB
 SEQ EEQIVFNENFPYLLRGSDSPKVLFFIELDFLSVDEIKNNSEVQNQECGFRKIAWAFK
 SEG
 COILS
 IgotB
 SEQ LLGANGNANINSKRLRLQLYPPPTKPSPLSVVEAFEWWSKCPRNHYPSTLYVTVRGLKVP
 SEG
 COILS
 IgotB
 SEQ DCIKPSYRSMALQEEKGKPVHCERHHSSSDTEPGLEESKEVIKWKRLPGQACRIPNK
 SEG
 COILS
 IgotB

```

SEQ      HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPILLYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECCCCCEEE

SEQ      SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFHPAVRELVTTCYDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEETTTEEEETT--TTCEEEETTTCETEEETT--TCEEEETT

SEQ      MIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEEETTTTBTTEEEEEECCCCC-EETEEETTEEEETTTEEEEEE.....

SEQ      EHSVHHWTINKEIKETEFKGIPISYLEIHPNGKRLLIHTKDSTLRIMDLRILVARKFVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLFAGSEDIYVYVNPETGEQVAMYSDFPKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAPGQNEPILLYIDFVHAQQAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FQIDFVHTSSSTKMQLVKQRLVTEVIRSCAAKVNKNLSFTSPPAVSSQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQELHQFGFTQTGTIISIERKPCNHQVDTAPTVALYDYTANRSDTLIHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPLSPPEETKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      OKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTMRKNKQAGRAKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

Prosites for DKFZphtes3_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005

PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

Pfam for DKFZphtes3_ln3.1

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrCHnnWVMCVaFSPDGrWFIvSGSWDgTCRLWD*		
Query	650	LCGHLNIIYDLWSKDDHY-ILTSSSDGTARIWK	682
HMM_NAME	Src homology domain 3		
HMM	*pyVIALYDYqAqdpDELSFkEGDIIiIEsDD.MWrgRnnnTNGQEGW		
Query	1054	PTVVALYDYTANRSELTIHrgDIrVFFKDNEDMWYGSIGK--GQEGY	1100
HMM	IPSNYVEPi*		
Query	1101	FPANHVASE	1109

DKFZphtes3_20c21

group: testes derived

DKFZphtes3_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
1 GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGGCGAGGT
51 ACGGGCCAGA AGTCGCAAGT ACCCAGCTGC TGCCACGTT TCTGGTCCAG
101 AGTCCCGAAC CCGGAGCACT GGGATGCTG GCTACTCGA GCCAAGGCAC
151 TGATGTTTGA ACTGGAACCT TCAAAACGTT TAATAAGAGT CTTCAGGATG
201 GGTTTGAAC TACACAAGTA GAAATTTCTT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTTTTGCTT TCCTGGAGAG GAGCTTGAAG AGGTGGTTCT
301 GCAGACAGCC ACAGTGATAC TCAGGAAACC AGAGGAATGG ATTTGACTTT
351 TCTGCTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTTGTA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACTTTCC
451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCAAGT CCAGATTAC TGGATTGGGG GATAGAGGAA CGGTACACAG
551 TAGGGAGAAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAGTCA
651 GCCTCGTGGT GGAATTAATT TTTCTTTAT GATGGTTCGA AGSTAAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTTC TTACTTTTAT CCTTCCAGA
751 CCCTGCTAGA CCAACAGGAG TTGCTTTGTG GACAGATTGC TGGAGTTGTC
801 CGCTGCTTT CTGCAATTTT TGACTCTGCT CACTACTTGT TTGCTCTGAG
851 AAAACGGAAG TTTGCCATAA AAGTTGATGG AGATTACCTT TGGGTGCTGG
901 CGTGTGCTGT GGAGCTCCCT GATGTCACT GCAAGCGGTT TCTGGATCAG
951 CTAGTTGGAT TCTTTAATT TTACAATGGA CCTGTTTCCC TAGCTTATGA
1001 GAATGTTTCT CAGGAAGAAG TGAGCAGCGA GTGGGACACC TTCATCGAGC
1051 AAATCTGTA AAACACCACT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAATAAATG GGAGCCCTGT TTGTTGCTGA AGGCAGCCCG
1151 CATTCTGAG ACCTGCCAGC GCTCGCCTCA CATTCTGCTT GGCTGCATCC
1201 TCTATAAAGC ACTGATTGTC AGCACCCAAC TCCCGCCCTC CCTCACCCGG
1251 AAGGTCCTGC TTCACCGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTACGCC CAGCACCATC CAAAGGGTGG GAGCAGTCT GCGCTGAAAG
1501 AAAAGCCAC TGCCCATGTT GAATCCATGG CCGGACACAC CCCAGATCCC
1551 ACATCCCTCT ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACTG CACAACCTGT
1651 CAGGGGTGGA GGTCTTGGC CTCAGCTCCT CCCTGGGAAA GGAACCTAGT
1701 TTTCTCAAG AAGAACTCGA CTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGAA ATGGCCTCAG GTCAATTTGC CTTCCTACAT GTGCCCTGTT
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCCAGCAGC
1851 CTGGAACCCA CGCTCCTTGA GGACACAGCC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTGGAAG
1951 ACCATCCTGG CCATAGCAGC CAAGCCCAAC TTCCAGAGGC AGACCCCTCT
2001 CCCAGAAGGA CCGCAGGCC CTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGAAAC AAGCTTCCCA CGGGGAAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGCCAC GCAGCCCTGT GTCTGGAATG CAGTTCAAGG
2151 TCAGCAAACT GTCAGGGTGC TGCCCTCTCT GCAGATGGAA TCAGCTCCAG
2201 GCTGACACCA GCAGAGTCTT GCATGGGGCT GTGAGGATG AATCTCTACA
2251 CTCAGTGGCT CAAAGGGCTG ATGCTGTCCC TGTGCTGTA GAGGCCGCTG
2301 CTGGAGACA CCGCAGCCAT AGAGGAAGTG TACCACACCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAGA GACGCTGCCC AGGGATGAGG
2401 CAGCTCCAC GAGCAGCACC TACAACCTCA CATATTACGA CCGCATTCAG
2451 AGCTTGCTGA TGGCAACCT GCCGAGGTG GCCACCCGCG ATGATCGCCG
2501 CTCTCTCAG GCGCTCAGCC TGATGCATAG CGAATTTGCC CAGCTGCCCG
2551 CGCTTTATGA AATGACTGTC AGAATGCTT CCACGGCTGT GTACGCCTGT
2601 TGCAACCCCA TCCAGGAGAC ATATTTCCAG CAGCTGGCAC CTGCAGCACG
2651 GAGCTCCGCC TTCCCAAAAC CTCAGGATGG CGCCTTCAGC CTCTCCGGCA
2701 AAGCAAAGCA GAAGTGTGCT AAGCAGCGGG TGAATTTGCT CTGAACCTGA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAGGAAG CTCTGCCTTT
2801 TATAATTGAA AAGGCCCTCT TATTTTATTT TTCTGAAA CATTCCCTTT
```

```
2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG GTTAGTCTTT CAGGTCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACTGTATT TAATCGGTAG TGTGGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTCATA TCTCCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCGAAGTT GGCATTTTTT
3101 CTAGAACCTG ATCGTCCACT AGCCCAGAGT GTGTGTGTTT AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGAGGT AGGTTTCTCC AGTGGGCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTTG AATTCGTGTT GAACTCTACT CAAATCTAGG GGCCTCTTCT
3351 TTGGACCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCCTG GCTGCCTTTG ACTCAAGTG GGCCTGTGAG
3501 CTGGTTCTTC TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGCTCTG
3551 GGGTAAAG CACTGTGCTT TTCAGTGGTG GCTGCGTAA AGGAGCGGAC
3601 ACTCAGCTGT GTGTTCTGGG GCTTGTGTGG TACTTAGAAC CTCAGTTCTA
3651 TTACGTATA GTCAGACATT TTTTTCAGAG TATGAGACAG ACTGCAGGAT
3701 GAAATATTT GTCAAAATCT TAACGAATG TTTACTGGAA GTACTTGAGA
3751 TTCCATTGGA GAGTTGTATT GTTAATAAT TCATGTCAGT GAACTGATAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS1048E9 from database EMBLNEW:
Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2
Contains pseudogene similar to ribosomal protein S3A and part of a gene
similar to C.elegans protein CE02118, ESTs, STS, GSS.
Score = 6540, P = 0.0e+00, identities = 1308/1308
-14 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708
Category: putative protein
Classification: no clue

```
1 MATSTSTEAK SASWMNYEFL YDGSKVKEEG DPTRAGICYF YPSQTLLDQO
51 ELLCGQIAGV VRCVSDISDS PPTIVRLRLK KFAIKVDGXY LWVLGCAVEL
101 PDVSKCRFLD QLVGFNFYV GPVSLAYENC SQEELSTWMD TFIEQILKNT
151 SDLHKIFNSL WNLQOTKVEP LLLLKAARIL QTCORSFIL AGCILYKGLI
201 VSTQLPPLST AKVLLHRTAP QEORLPTGGD APOEHGAALP PNVIIPVTV
251 TKEEAILLHE FPVEQMTSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWTTPO PTPDEACPD GRKENGCLSG HDLESIRPAG LHNARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPAEQEV EMASGHFAFL HVPVPDGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPPSA PEMLTQHGAG EQVEDHPGHS
451 SQAPIPRADP LPRRTTRRLL LPRLDPGQRG NKLPTEGQGL DEDVGVGCE
501 HAAPGLECSS GSANCOGAGP SADGISSRLT PAESCHMLVR MNLYTHCVKG
551 LMLSLLAEEP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTTYDRI QSLLMANLPQ VATPHDRRL QAVSLMHSEF AOLPALYEMT
651 VRNASTAVYA CCNPIQETYE QQLAPARSS GFNPFQDGAF SLGKAKQKXL
701 LKHGVNLL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_20c21, frame 3

Report for DKFZphtes3_20c21.3

[LENGTH] 708
[MW] 76900.23
[pI] 5.30
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.36 %

SEQ MATSTSTEAKSASWNNYFFLYDGSKVKEEGDPTRAGICYFYPSQTLDDQQLCGQIAGV
SEG .XXXXXXXXXXXXX.....
PRD ccc

SEQ VRCVSDISDSPPTLVRLRLKFAIKVDGYLWVLGCAVELPDVSKRFLDQLVGFFNFYN
SEG
PRD eeeeecc

SEQ GPVSLAYENCQEELSTEWDTFIEQILKNTSDLHKIFNSLWNLDTKVEPLLLKAARIL
SEG
PRD ccc

SEQ QTCQRSFILAGCILYKGLIVSTQLPPSLTAKVLLHRTAPQEQRLPTGGDAPQEHGAALP
SEG
PRD hhhcc

SEQ PNQIIPVFTKEEAI SLHEFPVEQMTSLAS PAGLDGSAQHHPKGGSTSAKENATGH
SEG
PRD ccc

SEQ VESMAWTTDPDTPDEACPDGRKENGCLSGHDLESIRPAGLHNSARGEVLGLSSSLGKEL
SEG
PRD ccc

SEQ VFLQEELDLSEIHIPEAQEVEMASGHFAFLHVPVPDGRAPYCKASLSASSSLEPTPPEDT
SEG
PRD hhhhhhcc

SEQ AISSLRPPSAPEMLTQGAQEQVEDHPGHSSQAPIPRADPLPRRTLRPLLPRLDPGQRG
SEG
PRD ccc

SEQ NKLPTGEQGLDEDVDGVCESHAAPGLECSSGSANCQAGPSADGISSRLTPAESCMGLVR
SEG
PRD ccc

SEQ MNLYTHCVKGLMLSLLAEPLLGDSAAIEEVYHSSLASLNGLEVHLKETLPDEAASTSS
SEG
PRD ceeeeehhhhhhhhhcc

SEQ TYNFTYYDRIQSLMANLPQVATPHDRRFLQAVSLMHSEFAQLPALYEMTVRNASTAVYA
SEG
PRD ccc

SEQ CCNPIQETYFQQLAPARSSGFNPQDGAFLSGKAKQKLLKHGVNLL
SEG
PRD ecc

(No Prosite data available for DKFZphtes3_20c21.3)

(No Pfam data available for DKFZphtes3_20c21.3)

DKFZphtes3_20k2

group: signal transduction

DKFZphtes3_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
101 GTGATGGAGA GTCTCTGCGG TGCCATCTGG GATGCAAAAC GTCCCTGTGT
151 CCCCACGCTC CAGGCCCTAG ATGCTCCCCG CCGGTCACTC ACTTACTCGT
201 CAGATCGCCC GTCTCGGTAT CACAGTGCCT CTGTTCAAGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGGC GACCCACTCC AAAAGGACAC TGCCCCAGAC CCCCTGGATG
351 GAGACCCCTAA CTCACGGCCA CTCACGCCA AGCCCCAGCT CTCACGGGCC
401 AAGAGCCGCA CCGGCTCTTT TGGGAAGGCT GACTCGGAGG AGGCTTTCCC
451 GGTGGATTGC CCTCAGGAGG AAGGTGAGCT GGAATCTGCG CCGACCATCA
501 CAGTCAGCCC TGTATCACC ATCCAGAGGC CAGGAGACGG CCCACCCGGT
551 GCCAGGCTGC TGTCCTCAGG CTCTGTCTGC GCCAGCACCG AGAAGACCTT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCCTTCTT CAGAATAACT
651 GCCAGGATCT CGAGAGCCTG CTGCTCTTCC TGCAGAGAGC CAAGAAGCAC
701 CTCACAGACA AGCAGTTCAA AGACCTCGAG ACAGGGAAGA CTTGTCTGCT
751 GAAAGCCATG CTCACCTGCG ATGACGGACA GAACACCACC ATCCCCCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
851 AGCTACACGG ACAGTACTA CAAGGGCCAG ACAGCAGTGC ACATCGCCAT
901 CGAGAGAGCG AACATGGCCC TGGTGACCTT CCTGGTGGAG AACGGAGCAG
951 ACGTCCAGGC TGGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTTG TCCTGGCCGG CGTGACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGCTGCA GAACTCTCTG CAGACGGCCG
1101 ACATCAGCGC CAGGACTCGG GTGGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCGG ACAACACGGC CGACAACACG AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCTTGG GGCCCAACT GCACCCGAGC CTGAAGCTGG
1251 AGGAGCTCAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGCTTT GGCCTATATT CTCACGGGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTCGCTGTAC GACCTGTCTT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAAT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCCT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTTGCCTCCC TTTAAGATGG AAAAAATTGG AGACTATTTT CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTTT CGAGGGATT
1751 CAGTATTTCC TGCAAGAGCG GCCCTGATG AAGACCTGTG TTGTGACAG
1801 CTACAGTGAG ATGCTTTTCT TTCTGAGTCT ACTGTTCTATG CTGGCCACCG
1851 TGGTGCTGTA CTTCAGCCAC CTCAGGAGAT ATGTGGCTTC CATGTTATTC
1901 TCCTTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTGTG TACATCGTCT TCTGTTTCGG GTTTTCCACA
2051 GCGGTGGTGA CGTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTGC CACAGGTGGC GGGGGCTGCG CTCAGGCCCC CCCGATAGCT
2151 CCTACACAGC CCGTACTGCC ACCTGCTGGG AGCTGTTCAA GTTCACCATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCTG CTGCTGGCCT ATGTAATICT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CCCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGGAG
2351 AGCAAGACA TCTGGAGCTT CGAGAGAGCC ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCCGCTCA GGCAGCTGTC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACCT GACCACCTGG AACACCACCG TGGGCATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGGCGT CAAGCGCACC CTGAGCTTCT
2601 CCCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```

```
2701 CGAGGAAGTT TATCTGCGAC AGTTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCCTTAG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACCAG TGCTCTGTCA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCACGC ATGTGCCCCA ATCTGTGCTG GACAAGCTGT GGGAAAGCGT
2951 CTGGGAAGCA TGGGCAAGTA TGTACATCA ACCGTCACTG TCCCCAAGTG
3001 AATCTCCTAA CAGACTTTCA GGTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGTGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTCCT TTGATTTTA
3101 TTCTTTTCTT TGAGACAGAA TTTCACCTCT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCTGCGCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGTCTGGCTA ATTTTGTGTA TTTTCTTAAT AGATATGGGG TTTCGCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATCTTTT TTTTCTTCT GTGAGACAGA GTTTCACCT
3451 TGTGTCACAG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCTTCCCGG GTTCAAGCGA TTCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGCT GAGCACTAC ACCCCCGGCT AATTTTGTGA TTTTAAATAG
3601 AGACGGGGTT TCACCATGTT GCCCAGGCTG GTCTCGAAGT CTGACCTCA
3651 GGTGATCTGC CCGCTTGGC CTCCCAAGT GCTGGGATTA CAGGTGTGAG
3701 CCGCTGCGCT CGGCTTCTT TGATTTTATA TTATTAGGAG CAAAAGTAAA
3751 TGAAGCCAG GAAACACCT TTGGGAACAA ACTCTCTCTT TGATGGAAAA
3801 TGCAGAGGCC CTCTCTCTCT GTGCGTGTCT TGCTCTCTCT ACCTGCCCGG
3851 GTGGTTTGGG GGTGTTGGTG TTCTCTCCTT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTTGGCA GAATCAAGT GTTGCAGCAG TGCTTCTCT
3951 ATCTCTCTT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAGATC TTCCATTTCT AGGAACCCCA GTCTGTGCTT TCCGCAATGG
4101 CACATGCTTC CACTCGATC ATACTGGCAT CCTCAAAATA ACAGATATGT
4151 ATACATATA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

99288727:
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:
A non-pungent triphenyl phenol of fungal origin, scutigerol, stimulates
rat dorsal root ganglion
neurons via interaction at vanilloid receptors.

Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839
Category: strong similarity to known protein
Classification: Cell signaling/communication

```
1 MKKWSSTDLG AAADPLQKDT CPDPLOGDPN SRPPFAKPQL STAKSRTRLF
51 KRGDSEEAFF VDCPHEGEL DSCPTTIVSP VITIQRFGD PTGARLLSQD
101 SVAASERTKL RLYDRRSIFE AVAQNWCOOL ESLLLFLQKS KKHLDNEFK
151 DPETGKTCIL KAMLNLDHGO NTIPLLEI ARQDLSKEL VNASVDSYY
201 KGQTAHIAI EERNMALVTL LVENGADVOA AAHGDFFKKT KGRPGFYFGE
251 LPLSLAECTN QLGIVKFLLO NSWOTADISA RDSVGNVLIH ALVEVADNTA
301 DNTKFVTSMY NEILILGAKL HPTLKLELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECRHLSRKF TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 NAAYYRFPVG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFLOS LEMLATVVLY FSHLKEYVAS MVFSLALQWT
551 NMLYYTRGFQ QMGIYAVNIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGNDSLPSSE STSHRWGSPA CRPPSSSYNS LYSTCLELFK FTIGMGDLEF
651 TENYDFKAVF IILLAYVIL TYILLNMLI ALMGETVNNKI AQESKNIWKL
701 QRAITLLOTE KSFLKCHRKA FRSGKLQVG YTPQCKODYR WCFRVDENVW
751 TWTWTVNGII NEDPGNCEGV KRTLSFSLRS SRVSCRHWKN FALVPLLEA
801 SARDRQSAQP EEVYLRQFSG SLKPDAEVF KSPAASGEK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20k2, frame 2

TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBLNEW:AB015231_1 product: "stretch-inhitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds.
Length = 838

HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query:      1 MKKMSSTDGAAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
M++ +S D  + P Q+++C DP D DPN +PPP KP + T +SRTLFGKGDSEEA P
Sbjct:      1 MEQRASLDSESESPPEQNSCLDPPORDPNCPPPVKPHITTRSRTRLFGKGDSEEAFF 60

Query:      61 VDCPHEEGELDSCTITVSPVITIQRPDGPFGARLLSQDSVAASSTKTLRLYDRRSIFE 120
+DCP+EEG L SCP ITVS V+TIQRPDGP R SQDSVA EK RLYDRRSIF+
Sbjct:      61 LDCPYEEGGLASCPITVSSVLTIQRPDGPASVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query:      121 AVAQNNCQDLESLLFLQKSKKHLTDNEFKDPETGKTCLLKAMLNLDGQNTTIPLLLEI 180
AVAQ+NCQ+LESLL FLQ+SKK LTD+EFKDPETGKTCLLKAMLNLDH+GQN TI LLL++
Sbjct:      120 AVAQSNCOELESLLPFLQSKKRLTDEFEKDPETGKTCLLKAMLNLDHNGQNTIALLLDV 179

Query:      181 ARQTDSLKELVNASYTDSSYKGTALHIAIERRNMLVTLVENGADVQAAAHGDFFKKT 240
AR+TDSLK+ VNASYTDSSYKGTALHIAIERRNM LVTLVENGADVQAAA+GDFFKKT
Sbjct:      180 ARKTDLSLKQFVNASYTDSSYKGTALHIAIERRNMLVTLVENGADVQAAAHGDFFKKT 239

Query:      241 KGRPGFYFGELPLSLAACTNQLGIVKFLQNSWOTADISARDSVGNVTVLHALVEADNTA 300
KGRPGFYFGELPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNVTVLHALVEADNT
Sbjct:      240 KGRPGFYFGELPLSLAACTNQLAIVKFLQNSWQADISARDSVGNVTVLHALVEADNTV 299

Query:      301 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLAALAGTGKIGVLAYILQREIQE 360
DNTKFVTSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct:      300 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKRGKLTPLAALASSGKIGVLAYILQREIHE 359

Query:      361 PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN 420
PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN
Sbjct:      360 PECRHLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN 419

Query:      421 RLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVVOGLPPFKMEK-IGDYFRVTGEI 479
RLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVVOGLPPFKMEK-IGDYFRVTGEI
Sbjct:      420 RLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVVOGLPPFKMEK-IGDYFRVTGEI 479

Query:      480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSLKEYVA 539
LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+QSLFML +VVLYFS KEYVA
Sbjct:      480 LSVSGGVYFFFRGIQYFLQRRPSLKSLEFVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query:      540 SMVFSALGWTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVYVFLFGFSTAVVTLI 599
SMVFSLA+GWTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVY+VFLFGFSTAVVTLI
Sbjct:      540 SMVFSLAWGTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVYVFLFGFSTAVVTLI 599

Query:      600 EDGKNDSLPESTSHRWGPACRPDSSYNSLYSTCLELFKFTIGMGDLFTENYDFKAV 659
EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCLELFKFTIGMGDLFTENYDFKAV
Sbjct:      600 EDGKNSLPESTPHKRGSAKCP-GNSYNSLYSTCLELFKFTIGMGDLFTENYDFKAV 658

Query:      660 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIKLQRAITILOTEKSFLKCMRK 719
FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIKLQRAITILOTEKSFLKCMRK
Sbjct:      659 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIKLQRAITILOTEKSFLKCMRK 718

Query:      720 AFRSGKLLQVGFTPDGKDDYRMCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR 779
AFRSGKLLQVG+TPDGKDDYRMCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR
Sbjct:      719 AFRSGKLLQVGFTPDGKDDYRMCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR 778

Query:      780 SSRVSGRHWKFNALVPLLRASARDROSAQPEEVYLRQPSGSLKPEDAQVFKSPAASGEK 839
S RVSGR+WKFNALVPLLR+AS RDR + Q EEV L+ ++GSLKPEDAQVFK SPAASGEK
Sbjct:      779 SGRVSGRHWKFNALVPLLRASARDROSAQPEEVYLRQPSGSLKPEDAQVFKSDMVPEK 838
```

Pedant information for DKFZphtes3_20k2, frame 2

Report for DKFZphtes3_20k2.2

(LENGTH) 839
(MW) 94950.75
(PI) 6.90
(HOMOL) TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus
vanilloid receptor subtype 1 mRNA, complete cds. 0.0
(FUNCAT) 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05
(PIRKW) alternative splicing 3e-06
(PIRKW) peripheral membrane protein 3e-06
(SUPFAM) ankyrin repeat homology 3e-06
(SUPFAM) unassigned ankyrin repeat proteins 3e-06
(PFAM) Ank repeat
(KW) TRANSMEMBRANE 4

SEQ MKKWSSTDLGAAADPLQKDTCPDPLDGPNSRPPPAKPOLSTAKSRTLFGKGDSEAF
PRD ccc
MEM
SEQ VDCPHEEGELDSCTPTITVSPVITIQPGOGPTGARLLSQDSVAATEKTLRLYDRRSIFE
PRD ccc
MEM
SEQ AVAQNQCQDLESLLFLQSKKHLTDNEFKDPETGKTCLLKLMLNLDGQNTTIPLLLEI
PRD hhhhechhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc
MEM
SEQ ARQDLSKELVNASYTDYKGTALHIAIERRNMLVTLVENGADVQAAAHGDFFKKT
PRD hhhcc
MEM
SEQ KGRPGFYFGEPLSLAACTNQLGIVKFLQNSWQTADISARSDVGNVTVLHALVEADNTA
PRD ccc
MEM
SEQ DNTKFTVSMYNEILILGAKLHPTLKEELTNKKGMTPLAAGTGKIGVLVILQREIQE
PRD chhhhhhhhhhhhhhhcc
MEM
SEQ PECRHLSRKFTWAYGPHSSLYDLSCIDTCEKNSVLEVIAYSSSETFNRHDMLLVEPLN
PRD ccc
MEM
SEQ RLLQDKWDRFVKRIFYFNLVLYCLYMIIFTMAAYRVPDGLPPFKMEKIGDYFRVTGEIL
PRD hhh
MEM
SEQ SVLGGVYFFFRGIQYFLQRRPSMKTLEVDYSEMLFFLOSLFMLATVVLYFSLKEYVAS
PRD ccc
MEM
SEQ MVFSLALGWTNMLYYTRGFQMGYAVMIEKMLRLDLCRFMFVYIVFLFGFSTAVVTLIE
PRD hhh
MEM
SEQ DGKNDLPSESTSHRWGPFACRPDSSYNSLYSTCLELFTIGMGDLEFFENYDFKAVF
PRD ccc
MEM
SEQ IILLAYVILTYILLNMLIALMGETVKNIAQESKNIWLQRAITILDTEKSLFKCMRKA
PRD hhh
MEM
SEQ FRSGKLLQVGYTPDGKDDYRWCFRVDEVNMTTNTNNGIINEDPGNCEGVKRTLFSLSRS
PRD hhcc
MEM
SEQ SRVSGRHWKNFALVPLLREASARDRQSAQPEEVYLRQFSGSLKPEDAEVFKSPAASGEK
PRD ccc
MEM

(No Prosite data available for DKFZphtes3_20k2.2)

Pfam for DKFZphtes3_20k2.2

WO 01/12659

PCT/IB00/01496

HMM_NAME	Ank repeat
HMM	*GyTPLHIAARYNNVEMVrLLQHGADIN*
	G+T+LHIA +++N+ +V LL+++GAD+
Query	202 GQTALHIAIERNNMALVTLLVENGADVQ 229

DKF2phtes3_2013

group: transmembrane protein

DKF2phtes3_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp

Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCTGCGTT GTTGTGAGT GGAGAGCAGG GAGTGGGGCC
51 AGCCAGCAGA AACAGTGGGC TGTACACAT CACCTTCAA TATGACAATT
101 GTACACACCA CTGGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCACG
151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCTTGAAA GGATTTCCGG
251 TAATACGGGA GGAGCTGAAG TCGGAGGGA GACAGTCCA ACAACTGATT
301 CTAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCTGAATA TGAATTTGA AACGATTAT TTCTGAAGG
401 TTGTCCCTTT TCCTTCCATT AAAACGAAA GCAATTACCA CCCTTCTTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGACCA ATCTAGCTTG
501 TAAACCCCTC TGGAAGCCTC GGAACCTGAA CATCAGCCAG CATGGCTCGG
551 ACATGCCAGT GTCCCTCGAC CAGCACCAGC ACAACTCTCG CTTCCTTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
701 TTTCTCCAGG GATATATATA ATTGAGCTGG TGGATGACRC TAACACAACA
751 AGAAAGTGA TCATTATGTC CTTAAAGCCA GTCCACTCCC CGTGGGCGGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGGTAGTC ATATCGGCAT
851 TCGGCACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAATATA
901 TATTACACAT TAGATGAAGA GAGCTCTGAG TCTTCCACAT ACACCTGCAGC
951 ACTCCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGGTCTTT CTCTGCTATT
1001 CCAGTAAAGA TGGCCAGAAAT CACATGAATG TCGTCCAGTG TTTGCGCTAC
1051 TTCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTAGCCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTCTATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGT CCAAGGTTAT GAACTCTTT
1201 GTGGACAAGA AGAATACAA ACACAAGGA GGTGGCCAG GCTCGGGGAA
1251 AGGAGAGCTC TTCTGTGTGG CGGTGTGAGC CATTCGCGAA AACCTCCGCC
1301 AGGCCAAGCA GAGTTCTGCC GCGGCGCTCA GCAAGTTTAT CGCGCTCTAC
1351 TTTGATTATT CTTGCGAGGG AGACGTGCCC GGTATCTTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCTCTA GCTCTGTTCC CACCTGCACCT
1451 CCCGAGACCA CGGCTCTCCG GAGCCGCGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAACCT ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACCTCGCCAT
1551 TTGCAACATG CACCAAGTTA TTGACGAGGA GCCCGACTGG TTGAAAAGC
1601 AGTTCTGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTTG ATTCTGGGCTT GGTTTTAAAT GATGTCATGT GCAAAACAGG
1701 CCTGAGAGGT GACTTCTGCC TAAAGTAGA GCGGCTCTT CTTGGGGCAA
1751 CCGGACACAG CGACTCCAG CACGAGAGTC AGCATGGGGC CTTGGACCAA
1801 GACGGGAGG CCGGCTCGC CTTGACGGT AGCGCCGCCC TGCAACCCCT
1851 GCTGCACACG GTGAAAGCCG GCAGCCCTC GGACATGCGG CGGAGCTCAG
1901 GCATCTATGA CTCGTCTGTG CCCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCCTCCTCT TCAGGCTCGG GTGAGGAGGA ACCTCCTGCC CTTCCTTCCA
2051 AGCTCCTCTC TTCTGGGTCA TGCAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGGCGT CGCCCTTTTG TAACAAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTASCTGC TGCCCTCCTC TGATTCGCCA GCTCATCTC
2201 CTGGTTGCTT GGCCTCCTG GAGCTGAGGT CTATACAGG GATATTTGGA
2251 GTGAATGCTT GGCCTGACTT TGTCTCCTT TGCCCAACCT CTTTACCGGA
2301 TATCTTGACA AACTCTCCAA TTTTCTAAA TGATATGGAG CTCTGAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2401 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595
Category: similarity to known protein
Classification: unclassified

```
1 MESQPFLNMK FETDYFVKVV PFPSIKNESN YHPFFFRTRA CDLLQPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKHGEPFKR
101 KTKQEQOTTE MTSCLLQNVN PGDYIILVDV DTNTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFV TLFVVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNMHNVV QCFAYFLQDF CGCEVALDLW
251 EDFSLCREGO REWVIQKINE SQFIIVVCSK GMRIFVDKRN YKHKGGGRGS
301 GKGEFLFLVAV SAIAELRQA RQSSSAALSR FIAYVFDYSC EGDVPGILD
351 STKYRLMDNL POLCSHLHSR DHGLQEPQQR TRQSGRRNYF RSKSGRSLYV
401 AICNMHQFID EEPDFWEKQF VPFHPPPLRY REPVLEKFDG GLVLNDVMCK
451 PGPESDFLCK VEAALVGATG PADSQHESQH GGLDQDGEAR PALDGSAAALQ
501 PLLHTVKAGS PSDMPRDSGI YDSSVPSSSL SLPLMEGLST DQTETSSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGSKADL GCRSYTDELH AVAPL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2013, frame 1

TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:NM31993.1 product: "interleukin 17 receptor"; Mus musculus
interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =
1.1e-13

>TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
mRNA, complete cds.
Length = 866

HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14
Identities = 85/284 (29%), Positives = 131/284 (46%)

```
Query: 213 KVFLCYSSKDGQNMHNVVQCFAYFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
KV++ YS+ D +++VV FA FL CG EVALDL E++ G WV QK +
Sbjct: 379 KVMIIYSA-DHPLYVDVVLKFAQLLTACGTEVALDLLEEQAISEAGVMTWVGRQKQEMV 437

Query: 269 HESQFIIVVCSKGMKY---FVDKKNYXXXXXXXXXXELFLVAVSAIAEXXXXXXXXX 324
+ IIV+CS+G + + + +LF A++ I
Sbjct: 438 ESNKIIIVLCRGRTRAKWQALLGRGAPVRLRCDHGKPGVGLFTAMNMHLPDFKRPACFG 497

Query: 325 XXXXXFIAYVF-DYSCGDPVPGILDSTKYRLMDNLPLQCSHLHSRDHGLQEPGQHTRO 383
++ YF + SC+GDPV + + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGTVPDLFGAAPRYPLMDRFEEV--YFRIQDLEMFQGRMHRV 550

Query: 384 G--SRNRYFRSKSGRSLYVAICNMHQFIDEEPWFEEKQFV---PFHPPPLR---YREP 434
G S NY RS GR L A+ + PDWE + + P L + EP+
Sbjct: 551 GELSGDNYLRSPGGRQLRAALDRFDWQVRCPDWFECENLYSADDQDAPSLDEEVFEEL 610

Query: 435 LEKFDGLVLNDVMCKPGPESDFLCKVEAALVGATGPADSQHESQHGLDQDGEAR 491
L +G+V + + P S CL ++ V G G A H L G+ P
Sbjct: 611 LPP-GTIVKRAPLVRE-PGSQACLAIDPLV-GEEGAATAKLEPH--LQPRGQPAP 662
```

Pedant information for DKF2phtes3_2013, frame 1

Report for DKF2phtes3_2013.1

[LENGTH] 595
[MW] 66847.05
[pI] 6.27
[HOMOL] TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus interleukin
17 receptor mRNA, complete cds. 2e-14
[BLOCKS] BL00740A MAM domain proteins
[BLOCKS] BL01224B N-acetyl-gamma-glutamyl-phosphate reductase proteins
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 13.61 %

SEQ MESQPFNLNMFETDYFVKVVPFSPKSNESNYHPFFRTRACDLLLQPDNLACKPFWKPRN
SEG
PRD ccc
MEM

SEQ LNISQHGSDMQVSFDHAPHNFGFRFFYLHYKLKHEGPFKRKTCKQEQTTEMTSCLLQNV
SEG
PRD eeeeecc
MEM

SEQ PGDYIIELVDDTNTTRKVMHYALKPVHSPWAGPIRAVAITVPLVVISAFATLFTVMCRKK
SEG
PRD ccc
MEMMMMMMMMMMMMMMMMM.....

SEQ QQENIYSHLDEESSESSTYTAALPRERLRPRPKVFLCYSSKDGQNHMNVVQCFAFLQDF
SEGxxxxxxx.....xxxxxxx.....
PRD hhhhhhhhhcc
MEM

SEQ CGCEVALDLWEDFSLCREGQREWVIQKIHESQFIIIVCSKGMKYFVDKKNYKHKGGGRGS
SEGxxxxxxx.....
PRD ccchhhhhhhcc
MEM

SEQ GKGELFLVAVSAIAEKLRAKQSSSAALSKEFIIVFDYSCGDVPGILDSTKYRLMDNL
SEGxxxxxxxxxxxxxxxx.....
PRD cccccceehhh
MEM

SEQ PQLCSHLHSRDHGLQEPGQHTRQGSRRNYFRSKSGRSLYVAICNMHQFIDEEDWFEKQF
SEG
PRD cchhhhhhhcc
MEM

SEQ VPFHPPPLRYREPVLKFDGLVLDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQH
SEG
PRD ecc
MEM

SEQ GGLDQDGEARPALDGSAAQLPLHTVKAGSPSDMPRDSGIYDSSVPSSLSLPLMEGLST
SEGxxxxxxxxxxxxxxxx.....
PRD ccc
MEM

SEQ DQTETSSLTESVSSSSGLGEEPPALPSKLLSSGSKADLGCRSYTDELHAVAPL
SEGxxxxxxxxxxxxxxxx.....
PRD hhhhhhhhhhecc
MEM

(No Prosite data available for DKFZphtes3_2013.1)

(No Pfam data available for DKFZphtes3_2013.1)

DKFZphtes3_20m18

group: nucleic acid management

DKFZphtes3_20m18 encodes a novel 132 amino acid protein with similarity to the *S. cerevisiae* mitochondrial carrier protein RIM2.

The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer. The RIM2/MRS12 gene encodes a predicted protein of 377 amino acids that is essential for mitochondrial DNA metabolism and proper cell growth. Inactivation of this gene causes the total loss of mitochondrial DNA and, compared to wild-type rho⁰ controls, a slow-growth phenotype on media containing glucose. The novel protein seems to be the human orthologue of this protein.

The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

similarity to carrier protein RIM2

Sequenced by MediGenomix

Locus: unknown

Insert length: 3572 bp

Poly A stretch at pos. 3530, polyadenylation signal at pos. 3510

```
1  GCCGCGGGGA GGGCTGTGCC GGTGCTTTC TGCAGCCGCA TCTCGGCCAG
51 CTCTCCTCGC CGTCCCCGGG GCGCTGTGCG TCTCCAGTCC GGGACCGAAG
101 CCGCCTGCCG TAGCGGGCGG CCAGATCCGC GTCCCCCTC AGCGGCCGGA
151 GGACATGCCG GAGAGAGAA TGAAGGAGG GACACGCTGG TGCATCTGTT
201 TGCCGGAGGA TGTGGTGGTA CAGTGGGAGC TATTCTGACA TGCCACTGCG
251 AAGTTGTAAA AACACGACTG CAGTCATCTT CTGTGACGCT TTATATTCTT
301 GAAGTTCAGC TGAACACCAT GGCTGGAGCC AGTGTCAACC GAGTAGTGTC
351 TCCCGGACCT CTTCAATGCC TAAAGGTGAT CTGGAAAAA GAAGGGCCTC
401 GTTCCTTGTT TAGAGGACTA GGCCCCAATT TAGTGGGGGT AGCCCCTTCC
451 AGAGCAATAT ACTTTGCTGC TTATTCAAAC TGCAAGGAAA AGTTGAATGA
501 TGTATTTGAT CCTGATTCTA CCCAAGTACA TATGATTTCA GCTGCAATGG
551 CAGGTATGAA TGTATAATAT TAAAAAATA AAAAACTTTC TGAACCTAG
601 AGGCTTAATA TTGAATTATA AGTTTGTAGT GAAAAATTGA TGATTAATGT
651 GCTTTTCTAT GATTAGATGA TTTTACGTT TATCGATATA AACCAAAATTA
701 GGTATATGTA AAATCTGTCA TCAGTTGACA TTTTGTAGT CAGGAGTTTA
751 CATGCTAGGG TACAAGTAAT ATATTTATAT TGCCTTGTGT AGTCCACTGA
801 ATGTTTAGTG ATCATTGTGA ACAGTTTAA GAATCCAACC ATAATTACAC
851 TATAAATAAG TTATGGAGCT GTAATTTACT CTCTCTCTCT CAATTTCTGT
901 TAGTGCCTTT TCCCTTTTGT CTGCATGTTT TGGCTTCTGT CTGAAATGTG
951 TCGGCAATTC TTGGTAAAGT ATTCATTTTG TCCTGTGCTC AAATGCTGAA
1001 ATTTTGTGTA GTGATGTATT ATTATTGACA ATTCAGTTAC TATGTGTATT
1051 TTTTAAATTT GTTTATTATT CTACATAATT CACACTAGAC AGCACCTGAA
1101 ATTTAGACAC TGGCTATGTG TACATGCTTA CTATAGAAAT GTTTCAGGA
1151 ACTCTCTGTT TCTGTCATCA CTGATAAGTA TATATGATTG TGAATTAATA
1201 TAACATAGTT TAGGTCTTTA CCCTGCCATA AAGATAAACA GTTGGTTTGA
1251 CCAATCTGGT TCTGGAATCA TTTGCTGCTA TGCATGTTAG ACAAAGCCAC
1301 GAACCTTGAT TTCCATTGTA AAATCTCTCC TAATATCTGA GATTTATTGT
1351 ATATTTACTC ATATCTCACA TTTTCAAAT ATGCTGTAAC TTTATAAACT
1401 GTAGCTGCTT TCATCAGCTA TTGATCAATA AATTGAATGT CAATTATGTG
1451 CTTAATAATG AGTGCCTTAA ACTGTTAAAC ACTTTTGGTT TAGAAATAAA
1501 GTGAATCAAT TTGACCTATA TACTTCATGA AGTAAGTAAG TTTGAAATAC
1551 AAATTTCTGA AAGGTCAATA GCCCTTATCG TATTACAAAT TGTTTTTAAG
1601 GCTTTTGTGA TTTATTAATT GTCAGTTGAT TCACTGAAGC TTTAAACTG
1651 GAAGGGACAA TCCAAGGTC AAAAGAGTGA AATACAATCA TTTACCAATA
1701 AGGAAACCTT GGGCAAATTA TGTAATTTAT GTGAACCTCT CTTAGCTTAC
1751 CCATGGAATG AGTCAAGTGG TCTACATAGA TTTGGATTTT GAGAATTAGT
1801 TCTTTTCATT AGTGTATATG AGATTATCTT GTTACAACCTA GAATTTATTT
1851 TAATGTAATT TTTACAGATG TTGAATATTA GTAGATAGGA TTTTCCCTC
1901 ACGAATTTGG ATGTAAGGTA AAGGTTGGTG GCCAGTGACA AACCTTATAA
1951 CCACCTTTAT AGGTTCTTTA AAAATATATT TGTGAATTAC CAGTGATTAT
2001 GTTTTGGGCT TATAACCTCA GATAATTATA AAGAAATGTT AATCTTATTT
2051 GAAAGAATTG GAATCTAGAA AGTTAGATGA GCAGTCATTT TATATTGATA
2101 TTTGTTATAT CAGTATAGCA AATGCAGAGG TTCAGAATAT CTTTATTTCC
2151 ACTGGAACAT CTTATTTTCT TAGAGTATCT CATCAGAATT TATTACTGTA
2201 TTTGTATCAC ATTGCAAGA ATTTCAGTAG AATTGTCAGT TTGCACTTTT
2251 TTCTCAAATG TGTACAAATG TTAACATATA GTTCATTTTT ATCTGTACAT
2301 TGATGCCAAT TCCCAACTTG AATTCCTCAA GTTTTGGTAA ACTTACAATC
2351 TCATACTTGT TCAGAGGTTA TTGCACTGTA CACTTACTGT GTAGAAAAAT
2401 CTGTTTGAAT TTGTTGTCAG TTACATTGTT CTGAGAACTG TGCTCTCAGA
2451 GCTTCTGTGC ACTATTCATG AGCATTAACA CTTAGCCTTG CAGTTTTTATA
```

```

2501 CATAACTATA TGGTTAGTAA AACTGAATGG TCCAATGCAG ACTCATTAAA
2551 GTAGGCTTTT GCCCCCTTTG TTCTTGAAAT AATCTAGACC AGATTACTCG
2601 GGGTTTTTTT TAGGATTATT TTTATAGGTC TAAATATGAA TGATTTGGGG
2651 GTATGAAGTA CTAAAGATA GTTCTGTGAA AAATCATTTT CAGCTGTCTA
2701 TTCAAGGGAA AAAATGCTAA CCTTGTCAC TTTACTACACA AAACCACT
2751 AAAATAAACC ATTAATGATA CTGCCTGCAA GATTTTAAAC CACCAGATAG
2801 CACACACATT AAGGATTAT AAGGCACTGT ACGTAATTT TATCCAAGT
2851 GACCTCTCAA TTCATTTTCA TTTTGCATT TATCCATATG AACTCATGTT
2901 TAATTTAGAT AATAAAAATT TATTTTATTA AAAGGACAGT TTATTTAAAG
2951 TGGGTCCTTT TATTGTGTGT AGTGCATACT ATAAGAATTT GTAAGCCTCT
3001 AAAGTTGAGC TATAAATTTT CATGCATTAA AAATTTGTTT CAGTTGTGAG
3051 GATATTTAAT CAGATTAAAT AATGTTGACT CTTAATATTT TGCTTGCTT
3101 TTTTCTTCTC TACACATGAC CTTTGACAGA CTAAGTATAT CTCAGCTATT
3151 GAGGGTATCT GTTTTGTGTC CTGTATATTT TGTTTAAATT AACTTGTATA
3201 TTCTTTGTA TACACCTAGG CACAGATGTA TGCAAAAAAA ATTTGTTAAA
3251 TTACTTCTTT CTTTATACTA ATTCTCAATT TTTAAAAGAT TTTATCTGGC
3301 ATGTATATAC TTTTATATAG AACATTATAA ATGTAAAGGA AATGAATCT
3351 AATTTTAAAT GGATTATGTA TTCATACAGT TATCTCAAT TTTTAAATA
3401 CTAATAATGT AATCATTGAA TGTTTCCTAC ATACGTAGTG GGTTTTATTT
3451 GCTCACAGCA TACAGTTATT TTTCAATTTA TGTTTTCTA TTAGACTTAA
3501 ATTCATTAT AATAAAGGCT TTTACTCATT AAATACAAA AAAAAAAA
3551 AAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

95198680:

Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria.

Peptide information for frame 1

ORF from 169 bp to 564 bp; peptide length: 132
 Category: similarity to known protein
 Classification: Intracellular transport and traffic
 Prosite motifs: LEUCINE_ZIPPER (27-49)
 MITOCH_CARRIER (26-36)

```

1 MSQRDLVHL FAGGCGGTVG AILTCPLEVV KTRLQSSSVT LYISEVQLNT
51 MAGASVNRV SPGPLHCLKV ILEKEGPRSL FRGLGPNLVG VAPSRAYFA
101 AYSNCKEKLN DVFDPDSTQV HMISAAMAGM NV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3_20ml8, frame 1

PIR:S44092 probable carrier protein c2 - *Caenorhabditis elegans*, N = 2,
 Score = 147, P = 1.5e-19

PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*), N = 1, Score = 230, P = 6.2e-19

>PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*)
 Length = 377

HSPs:

Score = 230 (34.5 bits), Expect = 6.2e-19, P = 6.2e-19
 Identities = 55/133 (41%), Positives = 80/133 (60%)

```

Query:      8 VHLFAGGCGGTVGAILTCPLEVVKTRLQSSS-VTLYISEVQLNTMAGA----SVNRVVSF 62
              VH AGG GG GA++TCP ++VKTRLQS + Y S+ +N G+ S+N V+
Sbjct:     54 VHFVAGGIGGMAGAVVTCPFDLVKTRLQSDIFLKAYSQA-VNISKGSTRPKSINYVIQA 112

```


Query: 63 GP-----LHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIFYAAYSNCCKLNDVFD--P 115
 G L + + ++EG RSLF+GLGPNLVGV P+R+I F Y K+ F+
 Sbjct: 113 GTHFKETLGIIGNVYKQEGFRSLFKGLGPNLVGVIPARSINFFTYGTTKDMYAKAFNNGQ 172

Query: 116 DSTQVHMISAAMAG 129
 ++ +H+++AA AG
 Sbjct: 173 ETPMIHLMAAATAG 186

Score = 77 (11.6 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 25/88 (28%), Positives = 39/88 (44%)

Query: 3 QRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVSP 62
 Q ++HL A G A T P+ ++KTR VQL+ SV + +
 Sbjct: 172 QETPMIHLMAAATAGWATATATNPILIKTR-----VQLDKAGKTSVRQYKNS 219

Query: 63 GPLHCLKVILEKEGPRSLFRGLGPNLVG 90
 CLK ++ EG L++GL + +G
 Sbjct: 220 WD--CLKSVIRNEGFTGLYKLSASYLG 245

Score = 71 (10.7 bits), Expect = 6.6e+00, P = 1.0e+00
 Identities = 28/91 (30%), Positives = 45/91 (49%)

Query: 12 AGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVSPGPLHCLKVI 71
 + G V +I T P EVV+TRL+ + + N G R + G + KVI
 Sbjct: 294 SAGLAKFVASIATYPHEVVRTRLRQTP-----KEN---G---KRKYT-GLVQSFKVI 338

Query: 72 LEKEGPRSLFRGLGPNLVGVAPSRAIFYAAY 102
 +++EG S++ GL P+L+ P+ I F +
 Sbjct: 339 IKEEGLFSMYSGLTPLHMRTPVNSIIMFGTW 369

Pedant information for DKFZphtes3_20m18, frame 1

Report for DKFZphtes3_20m18.1

[LENGTH] 132
 [MW] 13993.36
 [pI] 8.42
 [HOMOL] PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast (Saccharomyces cerevisiae) 7e-19
 [FUNCAT] 07.16 purine and pyrimidine transporters [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 08.04 mitochondrial transport [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 02.13 respiration [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 01.05.07 carbohydrate transport [S. cerevisiae, YPR021c] 3e-10
 [FUNCAT] 07.07 sugar and carbohydrate transporters [S. cerevisiae, YPR021c] 3e-10
 [FUNCAT] 07.99 other transport facilitators [S. cerevisiae, YEL006w] 1e-09
 [FUNCAT] 01.07.10 transport of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YIL006w] 3e-09
 [FUNCAT] 07.04.07 anion transporters (cl, so4, po4, etc.) [S. cerevisiae, YKL120w] 2e-08
 [FUNCAT] 01.03.19 nucleotide transport [S. cerevisiae, YPR011c] 3e-08
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKR052c] 4e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YJR095w] 2e-07
 [FUNCAT] 01.01.07 amino-acid transport [S. cerevisiae, YOR130c] 5e-05
 [FUNCAT] 07.10 amino-acid transporters [S. cerevisiae, YOR130c] 5e-05
 [FUNCAT] 01.04.07 phosphate transport [S. cerevisiae, YJR077c] 7e-05
 [FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YJR077c] 7e-05
 [BLOCKS] BL00215B Mitochondrial energy transfer proteins
 [BLOCKS] BL00215A Mitochondrial energy transfer proteins
 [PIRKW] duplication 6e-09
 [PIRKW] transmembrane protein 6e-09
 [PIRKW] mitochondrial inner membrane 4e-07
 [PIRKW] transport protein 5e-06
 [PIRKW] mitochondrion 7e-08
 [PIRKW] chloroplast 3e-08
 [SUPFAM] Btl protein 3e-08
 [SUPFAM] ADP,ATP carrier protein repeat homology 4e-09
 [SUPFAM] Caenorhabditis probable carrier protein c2 4e-09
 [SUPFAM] probable carrier protein YPR021c 6e-09
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MITOCH_CARRIER 1
 [PFAM] Mitochondrial carrier proteins
 [KW] Alpha_Beta

SEQ MSQRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV

PRD cccccceccccccccceccccchhhhhhhhhcccccccccccccccccccccc

SEQ SPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCCKELNDVFDPDSTQV

PRD cccchhhhhhhhhccccceccccceccccceccccceccccchhhhhhhhhcccccccccc

SEQ HMISAAMAGMNV

PRD chhhhhhcccc

Prosites for DKFZphtes3_20ml8.1

PS00029 27->49 LEUCINE_ZIPPER PDOC00029

PS00215 26->36 MITOCH_CARRIER PDOC00189

Pfam for DKFZphtes3_20ml8.1

HMM_NAME Mitochondrial carrier proteins

HMM *pFwkDFLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpR.....

Query 5 DTLVHLFAGGCGGTVGAILTCPLVVKTRLQSS-SVTLYISEVQLNTMA 52

HMMYkGMIdCFRwiwkNEGWRGLWRGLgANvIRYIPqWaIRFGFY

Query 53 GASVNRVVSPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY 102

HMM EFMKeMFiDyfgeddyWmWfwmnYMaGs*

Query 103 SNCKELNDVFDp-DSTQVHMISAAMAGM 130

DKFZphtes3_21d4

group: signal transduction

DKFZphtes3_21d4 encodes a novel 464 amino acid putative GTP exchanging factor related to RCC1.

RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator.

The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

similarity to RCC1-like G exchanging factor RLG

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="20"

Insert length: 2321 bp

Poly A stretch at pos. 2293, polyadenylation signal at pos. 2262

```
1 GGGTCACGCA AGATGGCGGC GCCCAGAGGC TGCTGAGGCG CGGAACGGAG
51 GATGGCGCTG GTGGCGTTGG TGGCTGGGGC TCGGCTGGGG CGGCGGCTGA
101 GCGGGCCGGG GCTGGGGCGA GGGCACTGGA CGGCGGCCAG GCGCTCCCGG
151 AGCCGGCCGC AAGCGGCAGA AGCCGAGGCG GAGGTGCCCG TGGTCCAGTA
201 CGTGGGCGAG CGCGCTGCCC GCGCCGATCG CGTCTTCGTG TGGGGCTTCA
251 GCTTCTCGGG GCGCTGGGC GTGCCTTCCT TTGTGGTGCC CAGCTCCGGG
301 CCCGGGCCCC GCGCCGGCGC CCGACCGCGC CGCAGGATCC AGCCCGTGCC
351 CTATCGCCTG GAGCTGGACC AAAAGATTTC ATCTGCTGCT TCGCGCTATG
401 GATTACACT GCTGTCTCTT AAGACTGCGG ATGTTACGAA AGTCTGGGGG
451 ATGGGACTCA ACAAGATTC TCAGCTTGA TTTACAGGA GCCGAAAGA
501 TAAACGAGG GGTACGAGT ATGTGTTGGA GCCCTCACC GTCTCCCTGC
551 CTCTGGACAG ACCTCAGGAG ACACGGGTGC TGCAGGTCTC CTGCGGCCGA
601 GCTCACTCTC TTGTGTTGAC TGACAGGGAA GGAGTCTTCA GCATGGGAAA
651 CAATTCTTAT GGGCAATGTG GAAGAAAGGT GGTGCGAAAT GAAATTTACA
701 GTGAAAGTCA CAGAGTCCAC AGGATGCAGG ACTTCGATGG CCAGTGGTTC
751 CAGGTGCGCT GTGGTCAGGA TCATAGTCTG TTCCTGACGG ATAAAGGAGA
801 AGTCTATTCT TGTGGATGGG GTGCTGATGG GCAAAACAGT CTGGGTCACT
851 ACAATATCAC CAGCTCGCCC ACCAAGCTGG GTGGAGACCT GCGGGGAGTG
901 AACGTTATCC AAGTTGCCAC CTACGGTGAT TGCTGCCTGG CCGTGTCCGC
951 CGACGGAGGA CTTTTTGGTT GGGGAACTC GGAGTACCTG CAGCTGGCCT
1001 CTGTCACTGA CTCACACAG GTGAATGTGC CCCGCTGCTT ACACCTCTCA
1051 GGAGTGGGGA AGGTGCGACA GGCTGCATGC GGTGGCACGG GCTGTGCAGT
1101 GTTAAACGGA GAAGGACATG TTTTGTCTG GGGCTATGGA ATTCTTGGGA
1151 AAGGTCCAAA CCTAGTGGAA AGTGCCGTCC CTGAAATGAT TCCACCCACT
1201 CTCTTTGGCT TGACGGAGTT CAACCCAGAA ATCCAGGTTT CCCGCATCCG
1251 ATGTGGACTC AGCCACTTTG CTGCACAGAC CAACAAAGGA GAGCTGTTTG
1301 TATGGGGCAA GAACATCCGA GGGTGCCTGG GAATCGGTGC CTTGGAGGAC
1351 CAGTATTTC CATGGAGGGT GACGATGCCT GGGGAGCCTG TGGACGTGGC
1401 ATGTGGCGTG GACCACATGG TGACCTGGG CAAGTCATTC ATCTAAACCT
1451 CCTTCACCTG CTTGGGCGGC CCCGTCCCGG GAACCACTGG CACTCCTTGG
1501 CAGAGGCCAG CGCGTGCCCA GCCCCCAGGG GTTCTTGAT GGTGGTGGCG
1551 GAGGACCCTG CGTGCACTGT GACGCTCTGT CCTGAATCCC TTAGCGGGTA
1601 CCTACCAGGA GGATCAGGGC AAGGTCCCTC TCCAGCTGCA GGTGAGGCCT
1651 GCGGAACATC GCTTGGATGG CAGCCTTTGG TGGGCCGCTG TGGCCCCGAC
1701 CTCTCTGTTT TCTCCAAGTA ACATGCGACG GTGTCTGGT TCACGTCTCG
1751 CTTGAGAAAG CCGTCTTAGG AAAGCTTAGC TTGAACACAG TGCTCGGGAG
1801 GTTCTGCTC TGTCTGTCAT GGCAGTCTCT TGGTTTGTGT CTGGCCAAGG
1851 CCATGCGTGT GCCTCGGACC GAGCCCCAGC TTAGGCGAGG GAGTCAGGCT
1901 GGTTCGGGCC CTCGGTTTTC ATTCAAGCCA CCCTGCTCAT GGCCCTTCCT
1951 GGCCTGCTGC CACACCGCAA GCTCGCTGGG GGGACACTAG AAGCACCGTG
2001 GCGTGGGATT CCATCTGGAG CTGTCCGCGA GCACAGCCCG CAGCCTCCCA
2051 CCACGCTCAC TGCTTGCTT GGAAGAGTTA AGAAGCCCTT CAGGAAGAGA
2101 ATCAGAGCTA AGTTCTCTG CGCCGAGGGC CCCGAGCATA TCCGCCAAGG
2151 CTCAGCTGCA GTGCCAGGCG GAGGAGGAAG ATCCAGAAAT TGTGAACAAT
2201 GTTTGATTTA GTAGCGTGAC TTGCCTTTCC CTTTAAAAAC ATCTTTTACA
2251 AATCTGTCTT GGAATAAAGT CTATTTTCTG CTTTTTGGTT TTTAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry HS203358 from database EMBL:
 human STS SHGC-31781.
 Score = 1748, P = 1.1e-72, identities = 376/394

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 52 bp to 1443 bp; peptide length: 464
 Category: similarity to known protein

```

1 MALVALVAGA RLGRRLSGPG LGRGHWTAAAR RSRSRREAAE AEAEPVQVQY
51 VGERAARADR VFWGFSFSG ALGVPSFVVP SSGPGPRAGA RPRRRIQVPV
101 YRLELDQKIS SAACGYGFTL LSSKTADVTK VWGMGLNKDS QLGFRHSRKD
151 KTRGYEYVLE PSPVSLPLDR PQETRVLQVS CGRAHSLVLT DREGVFSMGN
201 NSYGGQGRKV VENEIYSESH RVHRMQDFDG QVVQVACGQD HSLFLTDKGE
251 VYSCGWGADG QTGLGHYNIT SSPTKLGGDL AGVNVIVQVAT YGDCCLAUSA
301 DGGLEFGWNS EYLQLASVTD STQVNVPRCL HFSGVGKVRQ AACGGTGCAV
351 LNNEGHEVFW GYGILGKGPV LVESAVPEMI PPTLFLGLEF NPEIQVSRIR
401 CGLSHFAALT NKGELFVWVK NRGCLGIGR LEDQYFPWRV TMPGEPVDVA
451 CGVDHMTLA KSFI

```

BLASTP hits

Entry CEW09G3_5 from database TREMBLNEW:
 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3
 Score = 395, P = 9.3e-37, identities = 111/330, positives = 165/330

Entry Y032_HUMAN from database SWISSPROT:
 HYPOTHETICAL PROTEIN KIAA0032.
 Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry B38919 from database PIR:
 hypothetical protein 2 - human (fragment)
 Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry AF060219.1 from database TREMBLNEW:
 product: "RCC1-like G exchanging factor RLG"; Homo sapiens RCC1-like G
 exchanging factor RLG mRNA, complete cds.
 Score = 273, P = 4.0e-21, identities = 84/262, positives = 124/262

Entry S71752 from database PIR:
 giant protein p619 - human
 Score = 282, P = 1.1e-19, identities = 86/287, positives = 144/287

Alert BLASTP hits for DKFZphtes3_21d4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_21d4, frame 1

Report for DKFZphtes3_21d4.1

```

{LENGTH}      464
{MW}           49997.08
{pI}           8.74
{HOMOL}        TREMBL:CEW09G3_5 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3 5e-34

{FUNCAT}       04.07 rna transport [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       03.07 pheromone response, mating-type determination, sex-specific proteins
               [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       08.01 nuclear transport [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
               cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.01.04 rrna processing [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.03.03 trna processing [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       30.03 organization of cytoplasm [S. cerevisiae, YGL097w] 2e-09

```

```

(FUNCAT)      30.04 organization of cytoskeleton [S. cerevisiae, YAL020c] 4e-06
[BLOCKS]      BL00870I
[BLOCKS]      BL00625B Regulator of chromosome condensation (RCC1) proteins
[BLOCKS]      BL00625A Regulator of chromosome condensation (RCC1) proteins
[PIRKW]       blocked amino end 3e-16
[PIRKW]       nucleus 3e-16
[PIRKW]       duplication 4e-08
[PIRKW]       tandem repeat 3e-16
[PIRKW]       DNA binding 3e-16
[PIRKW]       mitosis 3e-16
[PIRKW]       leucine zipper 3e-21
[SUPFAM]      pheromone response pathway component SRM1 4e-08
[SUPFAM]      WD repeat homology 3e-21
[PROSITE]     MYRISTYL 7
[PROSITE]     RCC1_2_2
[PROSITE]     AMIDATION 2
[PROSITE]     CAMP_PHOSPHO_SITE 1
[PROSITE]     CK2_PHOSPHO_SITE 5
[PROSITE]     TYR_PHOSPHO_SITE 2
[PROSITE]     GLYCOSAMINOGLYCAN 3
[PROSITE]     PKC_PHOSPHO_SITE 7
[PROSITE]     ASN_GLYCOSYLATION 2
[PFAM]        Regulator of chromosome condensation (RCC1)
[KW]          All_Beta
[KW]          LOW_COMPLEXITY 13.58 %

```

```

SEQ  MALVALVAGARLGRRLSGPGLGRGHWTAARRSRSRREAAEAEVPPVVQYVGERAARADR
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhheeeccccccccchhhhhhhhhhhhhhhhhhhheeeehhhhhhhhh

```

```

SEQ  VFVWGFSEFSGALGVPSFVVPSSGPGPRAGARPRRRIQVPYRLELDQKISSAACGYGFTL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeccccccccccccccccccccccccccccccccccccchhhhhhhheeeccccceee

```

```

SEQ  LSSKTADVTKVWGMGLNKDSQLGFHRSRKDKTRGYEYVLEPSVPSLPLDRPQETRVLQVS
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eccccccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

```

SEQ  CGRAHSLVLTDRREGVFSMGNNSYGQCGRKVVENEIIYSESHRVHRMQDFDGQVVQVACGQD
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

```

SEQ  HSLFLTDKGEVYSCGWGADGQTGLGHYNITSSPTKLGGLAGVNVQVATYGDCCCLAVSA
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

```

SEQ  DGGLFGWGNSEYLQLASVTDSTQVNVPRCLHFSGVGKVRQAACGGTGCAVLNGEGHVFVW
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

```

SEQ  GYGILGKGNPLVESAVPEMIPPTLFGLTENPEIQVSRIKGLSHFAALTNKGELFVWGK
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

```

SEQ  NIRGCLGIGRLEDQYFPWRVTMPGEPVDVACGVDMVTLAKSFI
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccceee

```

Prosite for DKF2phtes3_21d4.1

PS00001	200->204	ASN_GLYCOSYLATION	PDOC00001
PS00001	268->272	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	82->86	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	333->337	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	14->18	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	219->222	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	290->294	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006

PS00007	209->217	TYR_PHOSPHO_SITE	PDOC00007
PS00007	208->217	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	238->244	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	302->308	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00009	12->16	AMIDATION	PDOC00009
PS00009	206->210	AMIDATION	PDOC00009
PS00626	179->190	RCC1_2	PDOC00544
PS00626	235->246	RCC1_2	PDOC00544

Pfam for DKFZphtes3_21d4.1

HMM_NAME	Regulator of chromosome condensation (RCC1)		
HMM	*IAAGqHHTVCLTqDGRVYtWG*		
	+A GQ+H++ LT++G VY++G		
Query	235	VACGQDHSFLFLTDKGEVYSCG	255

DKFZphtes3_21j15

group: transcription factors

DKFZphtes3_21j15 encodes a novel 898 amino acid protein with similarity human NY-CO-33 protein.

NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

strong similarity to "NY-CO-33"

complete cDNA, complete cds, potential start at bp 27, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 4407 bp

Poly A stretch at pos. 4321, polyadenylation signal at pos. 4301

```
1 CGCTGCAGCA GGTGTCACAG AGCCGCATGC TCCCGGAGCC CAGCCTCTTC
51 AGCACCGTGC AGCTGTACCG GCAGAGCAGC AAGCTCTATG GCTCCATCTT
101 CACGGGGGCC AGCAAGTTCC GCTGTAAGGA CTGCAGCGCT GCCTACGACA
151 CCCTGGTGGG GTTGACAGTG CACATGAACG AGACGGGGCA TTACCGCGAC
201 GACAACCATG AGACCGATAA CAACAACCCC AAGCGCTGGT CCAAGCCTCG
251 CAAACGCTCC TTGCTGGAAA TGGAAAGGAA GGAAGACGCC CAGAAGGTGC
301 TGAAGTGCAT GTACTGTGGC CACTCCTTTG AGTCCCTGCA GGATTGTAGT
351 GTCCATATGA TCAAAACAAA ACATAACCAA AAAGTGCCTC TGAAGGAACC
401 CGTCACTCCT GTCGCCGCCA AAATCATCCC TGCCACTCGG AAGAAAGCTT
451 CCCTGGAGCT GGAGCTCCCC AGTCCCCAG ATTCCACAGG TGGAAACCCC
501 AAAGCCACCA TCTCAGACAC CAACGATGCA CTTCAGAAGA ACTCCAACCC
551 TTACATCAGC CCAATAATC GGTACGGCCA CCAGAATGGG GCCAGCTATG
601 CATGGCACTT TGAGGCCCGG AAGTCGCAGA TCCTGAAGTG CATGGAGTGT
651 GGGAGCTCGC ATGACACCCT GCAGGAGCTC ACTGCCACA TGTGGTCAC
701 TGGCCACTTC ATCAAGGTCA CCAACTCTGC TATGAAAAAG GGAAGGCCCA
751 TTGTGGAGAC GCCTGTCACA CCTACCATCA CAACCTGCTT GGATGAGAAG
801 GTCCAGTCCG TGCCCTTGGC AGCCACCACC TTCAGTCCC CCTCCAATAC
851 ACCTGGCAGC ATCTCCCCAA AACTGAATGT GGAGGTCAAG AAGGAAGTCG
901 ACAAGGAGAA AGCGGTCACT GACGAGAAAC CTAAGCAAAA AGACAAGCCT
951 GGCGAAGAAG AGGAGAAGTG TGACATCTCT TCCAAATACC ATTACTTGAC
1001 TGAAAAATGAC TTAGAAGAGA GTCCCAAGGG GGGGCTTGAT ATCCTCAAAT
1051 CCTTGGAAAA CACAGTGACA TCCGCAATCA ACAAGGCCCA GAACGGCACT
1101 CCTAGCTGGG GGGGCTATCC CAGCATCCAT GCCGCTACC AACTTCCCAA
1151 CATGATGAAG TTGTCCCTGG GCTCGTCGGG GAAGAGCAGC CCCCTGAAAC
1201 CCATGTTTGG CAACAGTGAG ATTGTCTCCC CGACGAAAAA CCAGACCCCTG
1251 GTCTCTCCAC CCAGCAGCCA GACGTCCCCC ATGCCCAAGA CAAACTTTCA
1301 TGCCATGGAG GAGCTGGTGA AAAAGGTCAC TGAGAAAGTT GCCAAAGTGG
1351 AGGAGAAGAT GAAGGAGCCG GATGGGAAGC TTTCCCGGCC CAAGCGGGCC
1401 ACTCCCTCCC CATGTAGCAG CGAAGTCGGG GAACCCATCA AGATGGAGGC
1451 ATCCAGCGAT GGGGGCTTCC GCAGCCAGGA GAACAGCCCC AGCCCCCGGC
1501 GGGATGGGTG CAAGGATGGG AGCCCCCTCG CTGAGCCGGT GGAGAATGGC
1551 AAGGAGCTGG TGAAGCCCTT AGCCAGCAGT TTGAGTGGCA GCACGGCCAT
1601 CATCACCAGC CACCCGCTG AACAGCCTTT TGTAAACCTT TTGAGCGCCC
1651 TGCACTCAGT CATGAACATT CACCTGGGCA AGGCCGCCAA GCCCTCCCTG
1701 CCTGCCCTGG ACCCCATGAG CATGCTTTTC AAGATGAGCA ACAGCCTGGC
1751 GGAGAAGGCT GCTGTGGCCA CCCCGCCGCC CCTGCAGTCC AAGAAGGCAG
1801 ACCACCTCGA CCGCTATTTC TACCACGTCA ACAACGACCA GCCCATAGAC
1851 TTGACAAAAG GGAAGAGTGA CAAAGGCTGC TCCTTGGGTT CAGTGTCTCT
1901 GTCACCCACG TCCACAGCCC CGGCAACCTC CTCATCCACG GTGACAACGG
1951 CAAAGACATC TGCCGTCGTA TCATTATGT CAAACTCGCC GCTACGCGAG
2001 AATGCCTTGT CAGATATATC CGATATGCTG AAGAACTTGA CAGAGAGCCA
2051 CACGTCAAAA TCCTCCACTC CTTCCAGCAT CTCCGAGAAG TCTGACATTG
2101 ACGGGGCCAC TCTGGAGGAG GCTGAGGAGT CGACGCCCGC CCAGAAGAGG
2151 AAGGGCCGCC AGTCAAAC TGACCCCCAG CACCTCCTGA TCCTCCAGGC
2201 CCAAGTTTGC GCCAGCCTCC GGCAGACCTC AGAAGGGAAG TACATCATGT
2251 CAGACCTGAG CCCCAGGAG CGGATGCATA TCTCCAGGTT CACCGGGCTG
2301 CTCATGACCA CCATCAGCCA CTGGTGGGCC AACGTGAAAT ACCAGCTTCG
2351 AAGGACAGGT GGAACAAAGT TCCTCAAAAA CTTGGACACT GGCCACCCCG
2401 TCTTCTTTTG TAACGATTGT GCGTCCCAAA TCAGGACTCC TTCCACGTAC
2451 ATCAGTCACC TAGAGTCACA CTTAGGCTTC CGGCTACGGG ACTTATCCAA
2501 ACTGTCCACC GAACAGATTA ACAGTCAGAT AGCACAAACC AAGTCACCGT
2551 CAGAAAAAAT GGTGACGTCC TCCCCGAGG AAGACCTGGG GACTTCCTAT
2601 CAGTGCAAAC TTTGCAATCG GACCTTGGC AGCAAGCAGC CTGTTAAACT
```

```

2651 TCACCTTAGC AAAACACACG GGAATCTCC GGAAGACCAC CTTCTGTATG
2701 TCTCTGAGTT AGAGAAGCAG TAGCATTGTC TTTTGATAGA AAGGACTGCA
2751 GTTTGCTTTG AGGGAACTG TGGAAAGCAC CTTCAGGCC CCTCTGACTT
2801 GTTGTCTTGC GCACATGTTT TTATTTTAAC TGCAGAGAAT CACTCTGGGC
2851 TGGACTGTTT TGTATAACTG TACAGTGTTC AATAGAGGTG CATAATCAGC
2901 TGTTGTACTT GGTAAAAATAT GAAGGTAAAA ATGCAGTGGT AAGTGTGTTG
2951 AACTTTGTGT AAACGGGATT TAGTTGTGAG CATCCTCCCG ATGCTTCAAG
3001 CTGCATGCAT TAACAGACAG TTTAATTAAG CATTATAAC GGAATCAGGC
3051 ACACCTTTTC CACGAGACTC GAGTGTGCTG GCATTTCTCA CCTTTTCATC
3101 TTTAGCCCTC TGAGTACTTT GAAGCACTTT TGCATTAATT TGGTAAAAAA
3151 ATAAAAATAA ATAATAATAA TGTATGAAGC TCTGTTTTT AACTCCTTA
3201 CCAGCTTAGT TATAATGAAT AATATGAACC TCCATTATG CAGGTCTGCA
3251 GGGGTATAAC ACGCCTTGAA ATTTAAAGA ATATTATTT CACATTGAAA
3301 CATAGATGTA TATATTGTAT AGATTTCAGA CTCTCTTATG AAAAAAATG
3351 TGATTGTGGT TAAATGACCT TTTCTTGCA TTTATAGCAA CAGTGTTTTA
3401 TGCACCTGCT ATGCTCTGGG CATAAGCTGT GCCTATGTAT AGTGTATATT
3451 TCTTTTTTTC TTTTTTTTAA GGTCTATGGG TTTGTTTTT TACATGCAA
3501 CATTGTAAAT TATACAGAAG ATACCACAGA TAGCATTAT AAAGTATACA
3551 GAAACATTAT CTGAAAGCAA AGTATGATAG TTTGTTTTGC TATACAGTAC
3601 ATCTATATTG ATAGAGGTTT ATGTTTAAAT TATACATATT TATTAGCATC
3651 ATATTGTGAT TTGTTTTGAG CAGTCTGAAT AAACGAGACC GGGAAAGACA
3701 CCCCTGGCAG GCATCAGAAC TATTTTGAC ATGATTTTA AAGGTATTTA
3751 TTAGAAATCA AAGAACACTC AAAATAAACT CAGTGCTCAA AGGGTTAAGT
3801 CTATTTGAAA AGGTTAAAAA AAAGAACAAA AAAAAAATAA GAACTTGATC
3851 TGTATTTTCT AAACATTGAT AAAGCCTTTA AAATGTTTGT ACTGTAATAC
3901 TTTGCTTAAA AGTCATGAGG CATTCTGTGA TCCAACCTCT TTCACTTATT
3951 TATAAGCCTT CTTGGTTGCT ATTCCATATT GTAGGATGCC TTTCTATTTT
4001 AATTGGTAACT TTTCTGTTTT GTTCTTCTTA ATTATTCTCC CAAGATCCCA
4051 CACTCGACGT TTATCTTTAG GCTTATGAAA GGTAAACCCGT GGTTACCGGC
4101 TCTCCAAAGT ATTCTGTTCT TCTCCATTTT TGGCAGTTAA TTTGCAGAAG
4151 TAACTGACAG CTGACACCAT ATGAGAACCT TTGTATAAAA TATTGGCATG
4201 TAAACAGCAC AGACACCGTA ACACACTCTG TGCCCTGTTT GGTGTTGAC
4251 AATGAAGCAC CATTATGTGA CTCTTCATAT AACCCTTTTT TCTACGGCAG
4301 CATTAAAAAT GTCTTTTTCG TATAAAAAAA AAAAAAATAA AAAAAAATAA
4351 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4401 AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 2720 bp; peptide length: 898
 Category: strong similarity to known protein

```

1 MLPEPSLFST VQLYRQSSKL YGSIFTGASK FRCKDCSAAY DTLVELTVHM
51 NETGHYRDDN HETDNNPNKR WSKPRKRSLL EMEGKEDAOK VLKCMYCGHS
101 FESLQDLNVH MIKTKHYQKV PLKEPVPVPA AKIIPATRK ASLELELPSS
151 PDSTGGTPKA TISDTNDALQ KNSNPYITPN NRYGHQNGAS YAWHFEARKS
201 QILKCMCEGS SHDTLQELTA HMMVTGHFIK VTNSAMKKKG PIVETPVTPT
251 ITTLLDEKQV SVPLAATTFT SPSNTPASIS PKLNVEVKKE VDKEKAVTDE
301 KPKQKDKPGE EEKCDISSK YHYLTENDLE ESPKGGDLIL KSLENTVTSA
351 INKAQNGTFS WGGYPSIHAA YQLPNMMKLS LGSSGKSTPL KPMFGNSEIV
401 SPTKNQTLVS PPSSQTSMP KTNFHAMEEL VKKVTEKVA VEEKMKEPDG
451 KLSPPKRATP SPCSSEVGEP IKMEASSDGG FRSQENSPSP PROGCKDGSF
501 LAEPVENGKE LVKPLASSLS GSTAITDHP PEQPFVNPLS ALQSVMIHL
551 GKAAPKSLPA LDPMSMLFKM SNSLAEKAAV ATPPLQSKK ADHLDRYFYH
601 VNNQPIDLT KGKSDKGCSL GSVLLSPTST APATSSSTVT TAKTSAVVSF
651 MSNSPLRENA LSDISDMLKN LTESHTSKSS TPSSISEKSD IDGATLEEAE
701 ESTPAQKRKG RQSNWNPOHL LILQAQFAAS LRQTSEGKYI MSDLSPOERM
751 HISRETFGLSM TTISHWLANV KYQLRRGTGG KFLKNLDTGH PVFFCNDCAS
801 QIRTPSTYIS HLESHLGFERL RDLKSLSTEQ INSQIAQTKS PSEKMTSSP
851 EEDLGTSYQC KLCNRTFASK HAVKHLHLSK HGSKSPDHLL YVSELEKQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21j15, frame 3

TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds., N = 1, Score = 1039, P = 5.5e-105

PIR:A38437 probable homeotic protein tsh - fruit fly (Drosophila melanogaster), N = 3, Score = 158, P = 7.2e-09

TREMBL:CE33058_1 gene: "unc-89"; product: "UNC-89"; Caenorhabditis elegans UNC-89 (unc-89) gene, complete cds., N = 2, Score = 175, P = 3.3e-07

>TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.
Length = 687

HSPs:

Score = 1039 (155.9 bits), Expect = 5.5e-105, P = 5.5e-105
Identities = 244/504 (48%), Positives = 319/504 (63%)

Query: 170 QKNSNPYITPNNRYGHQNGASYAWHFARKSQILKCEGSSHDTLQELTAHMMVTGHFI 229
QK +NPY+TPNNRYG+QNGASY W FEARK+QILKCEGSSHDTLQ+LTAHMMVTGHF+
Sbjct: 14 QKAANPYVTPNNRYGQNGASYTWFARKAQILKCEGSSHDTLQQLTAHMMVTGHFL 73

Query: 230 KVTNSAMKKGKPIVETPVTITTLDEKVQSVPLAATTFTS-PSNT----PASISPKLN 284
KVT SA KKGK +V PV ++EK+QS+PL TT T P+++ P S +
Sbjct: 74 KVTTSASKKQQLVLDPV-----VEEKIQSIPLPPTTHTRLPASSIKKQPDSPAGSTT 126

Query: 285 VEVKKEVDKEKA-VTDEKPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGLDILKSL 343
E KKE +KEK V + K K++ + EK + S+ Y YL E DL++SPKGGLDILKSL
Sbjct: 127 SEEKKEPEKEKPPVAGDAEKIKEESEDSEKFEFSTLYPYLREEDLDDSPKGGLDILKSL 186

Query: 344 ENTVTSAINKAQNGTPSWGYPYPSIHAAYQLPMMKLSLGSSGKSTPLKPMF-GNSEIVSP 402
ENTV++AI+KAQNG PSWGGYPYPSIHAAYQLP +K L ++ +S ++P + G + +S
Sbjct: 187 ENTVSTAISKAQNGAPSWGYPYPSIHAAYQLPGTVK-PLPAVQSVQVQPSYAGGVKSLSS 245

Query: 403 TKNQTLVSPSSQTSMPKTNFAMEELVKVTEKV-AKVEEKMKEPDGKLSPPKRATPS 461
++ L+ P S T P K+N AMEELV+KVT KV K EE+ E + K S K A S
Sbjct: 246 AEHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNKKEERPEKE-KSSLAKAA--S 302

Query: 462 PCSSEVGEPKMEASSDGGFRSQENSPSPRDGCKDGSPLAEPVENGKELVKPLASSLSG 521
P + E + K E S + Q+ P K PL NG E +K ++
Sbjct: 303 PIAKENKDFPKTEEVSG---KPQKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCN 359

Query: 522 STAIITDHPPEQPFVNPLSALQSVNMNIHLGKAAKPSLPALDPMMLFKMNSLAEKAABA 581
+ II DH PE F+NPLSALQS+MN HLK +K P +LDP++ML+K+SNS+ +K
Sbjct: 360 NLGIIMDHSPSPFVNPLSALQSIMNTHLGKVSQVSPSLDPLAMLYKISNSMLDKPVYP 419

Query: 582 TPPPLQSKKADHLDYFYHVNNQPIDLTGKSDK-GCSLGSVLLSPTSTAPATSSSTVT 640
P K+AD +DRY+Y N+DQPIDLT K+ S+ + SP + S +
Sbjct: 420 ATPV---KQADAIDRYYE-NSDQPIDLTGKSKNPLVSSVADSVASPLRESALMDISDMV 475

Query: 641 TAKTSAVVSFMSN-SPLRENALSDISDMLKNLTE 673
T + S S + E + +D S + L E
Sbjct: 476 KNLTRGLTPKSSPTSTVSEKSDADGSSFEALDE 509

Score = 865 (129.8 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 211/434 (48%), Positives = 268/434 (61%)

Query: 447 EPDGKLSPPKRATPSPCSSEVG--EPIKMEASSDGGFRSQENSPSPRDG-CKDGSPLAE 503
E + L P TP P S V E + + + + +E P + K SP+A+
Sbjct: 247 EHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNKKEERPEKEKSSLAKAASPIAK 306

Query: 504 -----P-VE--NGKELVK-PLASSLSGSTAIITD-HPPE--QPFVNPLSALQSVNMNIHLG 551
P E +GK K P A + D H P +P ++ + + I +
Sbjct: 307 ENKDFPKTEEVSGKPKQKGEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCENNLGIIMD 366

Query: 552 KAAKPSLPALDPMMLFKMNSLAEKAABATPPPLQSKKADHLDYFYHVNN---DQPID 608
+ +PS ++P+S L + N+ K + P L D L Y ++N D+P+
Sbjct: 367 HSPEPSF--INPLSALQSIMNTHLGKVSQVSPSL-----DPL-AMLYKISNSMLDKPV- 417

Query: 609 LTKGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENALSDISDML 668
K S P + + S+V ++ SPLRE+AL DISDM+
Sbjct: 418 -YPATPVKQADAIDRYYYENSQPIDLTGKSKNPLVSSVADSVASPLRESALMDISDMV 475

Query: 669 KNLTESHTSKSSTPSSISEKSDIDGATLEEA-EESTPAQKRKGRQSNWNPOHLLILQAQF 727
KNLT T KSSTPS++SEKSD DG++ EEA +E +P KRKGRQSNWNPOHLLILQAQF

Sbjct: 476 KNLTGRLTPKSSTPSTVSEKSDADGSSFEALDELSPVHKRKRQSNWNPQHLLILQAQF 535

Query: 728 AASLRQTSEGGYIMSDLSPOERMHISRFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 787
A+SLR+T+EGYIMSDL PQER+HIS+FTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD

Sbjct: 536 ASSLRRETTEGGYIMSDLGPOERVHISKFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 595

Query: 788 TGHVPVFFCNDCASQIRTPSTYISHLESHLGFRLRDLKSLSTEQINSQIAQTKSPSEKMV- 846
TGHVPVFFCNDCASQ RT STYISHLE+HLGF L+DLSKL QI Q +K + K +

Sbjct: 596 TGHVPVFFCNDCASQFRTASTYISHLETHLGFSLKDLKSLPLNQIQEQQNVSKVLTNKTG 655

Query: 847 -TSSPEEDLGTSYQCKLCNRTFASK 870
+ EEDLG+++QCKLCNRTFA +

Sbjct: 656 PLGATEEDLGSTFQCKLCNRTFAKQ 680

Score = 98 (14.7 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 32/95 (33%), Positives = 47/95 (49%)

Query: 90 KVLKCMYCGHSFESLQDLSVHMIKTKHYQKVPL-----KEPVT-PVAAKIIPATRKAS 142
++LKCM CG S ++LQ L+ HM+ T H+ KV K+ V PV + I + +

Sbjct: 45 QILKMECGSSHDTLQQLTAHMMVTGHFLKVTTSASKKGKQLVLDPVVEEKIQSIPLPPT 104

Query: 143 LELELPSS-----PDSTGGTPKATISDTNDALQKNSNP 175
LP+S PDS G+ T S+ +K P

Sbjct: 105 THTRLPASSIKKQPDSPAGS---TTSEEKKEPEKEKPP 139

Score = 81 (12.2 bits), Expect = 4.6e-93, P = 4.6e-93
Identities = 13/29 (44%), Positives = 20/29 (68%)

Query: 28 ASKFRCKDCSAAYDTLVELTVHMMNETGHY 56
A +C +C +++DTL +LT HM TGH+

Sbjct: 44 AQILKMECGSSHDTLQQLTAHMMVTGHF 72

Pedant information for DKFZphtes3_21j15, frame 3

Report for DKFZphtes3_21j15.3

[LENGTH] 898
[MW] 98486.72
[pI] 8.61
[HOMOL] TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens
antigen NY-CO-33 (NY-CO-33) mRNA, complete cds. 0.0
[BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins
[PIRKW] zinc finger le-06
[PIRKW] DNA binding le-06
[PIRKW] transcription regulation le-06
[PROSITE] MYRISTYL 9
[PROSITE] ZINC_FINGER_C2H2 4
[PROSITE] CAMP_PHOSPHO_SITE 5
[PROSITE] CK2_PHOSPHO_SITE 19
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 4
[PFAM] Zinc finger, C2H2 type
[KW] Alpha Beta
[KW] LOW_COMPLEXITY 11.36 %

SEQ MLPEPSLFSTVQLYRQSSKLYGSIFTGASKFRCKDCSAAYDTLVELTVHMMNETGHYRDDN
SEG
PRD ccc

SEQ HETDNNNPKRWSKPRKRSLLEMEGKEDAQKVLKCMYCGHSFESLQDLSVHMIKTKHYQKV
SEG
PRD ccc

SEQ PLKEPVTVAAKIIPATRKASLELELPSSPDSTGGTPKATISDTNDALQKNSNPYITPN
SEG
PRD ccc

SEQ NRYGHQNGASYAWHFEARKSQILKMECGSSHDTLQELTAHMMVTGHFIKVTNSAMKKGK
SEG
PRD ccc

SEQ PIVETPVTPTITLLDEKVQSVPLAATFTFTSPSNTPASISPKLNVEVKKEVDKEKAVTDE
SEG
PRD ccc

SEQ KPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGLDILKSLENTVTSAINKAQNGTPS

```

SEG      x.....
PRD      cccccccccccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhcccccc

SEG      WGGYPSIAHAYQLPNNMKLSLGSSGKSTPLKPMFGNSEIVSPTKNQTLVSPSSQTSMPM
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEG      KTNFHAMEELVKKVTEKVAKEEKMKEPDGKLSPPKRATPSPCSSEVGEPIKMEASSDGG
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      ccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccc

SEG      FRSQENSPSPPRDCKDGSPLAEPVENGKELVKPLASSLSGSTAIITDHPPEQPFVNPLS
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEG      ALQSVMMNIHLGKAAPSLPALDPMSMLFKMSNSLAEKAAVATPPPLQSKKADHLDRYFYH
SEG      .....
PRD      chhhhhhhccccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccc

SEG      VNNDQPIDLTGKSGDKCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENA
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      eccccceeeccccccccccccceccccccccccccceeeceeeeeeccccccccchhh

SEG      LSDISDMLKNLTESHTSKSSTPSSIIEKSDIDGATLEAEESTPAQKRKGRQSNWNPOHL
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      hhhhhhhhhhhhhccccccccccccceccccccccchhhhhhhhhccccccccchhh

SEG      LILQAQFAASLRQTSEKGYIMSDLSPQERMHISRFTGLSMTTISHWLANVKYQLRRTGGT
SEG      .....
PRD      hhhhhhhhhhhhhccccceeeccccccccchhhhhhhccccchhhhhhhhhhhhhhhcccc

SEG      KFLKNLDTGHPVFVFCNDCASQIRTPSTYISHLESHLGFRLRDL SKLSTEQINSQIAQTKS
SEG      .....
PRD      ceccccccccceccccceeeccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhcc

SEG      PSEKMVTSSPEEDLGTSYQCKLCNRTFASKHAVKLHL SKTHGKSPEDHLLYVSELEKQ
SEG      .....
PRD      ccceeeccccccccceehhhhhhhhhhhhhhhhhccccccccccccceeeeeecccc

```

Prosites for DK72phtes3_21j15.3

PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	405->409	ASN_GLYCOSYLATION	PDOC00001
PS00001	670->674	ASN_GLYCOSYLATION	PDOC00001
PS00001	864->868	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	75->79	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	139->143	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	432->436	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	456->460	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	157->160	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00005	641->644	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00005	686->689	PKC_PHOSPHO_SITE	PDOC00005
PS00005	730->733	PKC_PHOSPHO_SITE	PDOC00005
PS00005	842->845	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	78->82	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	161->165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	684->688	CK2_PHOSPHO_SITE	PDOC00006
PS00006	689->693	CK2_PHOSPHO_SITE	PDOC00006
PS00006	695->699	CK2_PHOSPHO_SITE	PDOC00006
PS00006	745->749	CK2_PHOSPHO_SITE	PDOC00006

PS00006	810->814	CK2_PHOSPHO_SITE	PDOC00006
PS00006	840->844	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	884->888	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	732->740	TYR_PHOSPHO_SITE	PDOC00007
PS00007	883->892	TYR_PHOSPHO_SITE	PDOC00007
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	156->162	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	617->623	MYRISTYL	PDOC00008
PS00008	757->763	MYRISTYL	PDOC00008
PS00028	795->816	ZINC_FINGER_C2H2	PDOC00028
PS00028	860->882	ZINC_FINGER_C2H2	PDOC00028
PS00028	33->56	ZINC_FINGER_C2H2	PDOC00028
PS00028	94->117	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKF2phtes3_21j15.3

HMM_NAME Zinc finger, C2H2 type

HMM *CpwPDCgKtFrrwsNlrRHMR..T.H*
 C++ C ++ + +L+ HM+ H
 Query 33 CKD--CSAAYDTLVELTVHMMNET-GH 55

26.69 (bits) f: 94 t: 116 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNlrRHMR..T.H*
 C + CG +F + +L HM+ H
 dkfzphes3 94 CMY--CGHSFESLQDLSVHMIKT-KH 116

Query f: 795 t: 815 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

HMM *CpwPDCgKtFrrwsNlrRHMRTH*
 C++ C R++S+++ H+ +H
 Query 795 CND--CASQIRTPSTYISHLESH 815

27.12 (bits) f: 860 t: 881 Target: dkfzphes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNlrRHMR.T.H*
 C+ C++TF +++ + H+ H
 dkfzphes3 860 CKL--CNRTFASKHAVKLHLSK-TH 881

DKFZphtes3_21116

group: intracellular transport and trafficking

DKFZphtes3_21116 encodes a novel 66 amino acid protein nearly identical to rat ribosome attached membrane protein 4 (ramp4).

The novel protein seems to be the human orthologue of rat ramp 4. Ramp4 is involved in the regulation of translocation of proteins into endoplasmic reticulum, e.g. of the MHC class II associated invariant (gamma) chain.

The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

identical to rat ribosome attached membrane protein 4

ORF Bp 316-513 (66 aa) see BLASTX

Sequenced by LMU

Locus: unknown

Insert length: 2488 bp

Poly A stretch at pos. 2464, polyadenylation signal at pos. 2442

```
1 CTTCTCTTT CACTCCGCGC TCACGGCGGC GGCCAAAGCG GCGGCGACGG
51 CCGCGCGAGA ACGACCCGGC GGCCAGTTCT CTTCTCTCTG CGCACCTGCC
101 CCGCTCGGTC AGTCAGTCGG CGGCCGCGCG CCGGCTTGTG CTCAGACCTC
151 GCGCTTGCGG CGCCAGGCC CAGCGGCCGT AGCTAGCGTC TGGCCTGAGA
201 ACCTCGGCGC TCCGGCGCGC CGGGCACCAC GAGCCGAGCC TCGCAGCGGC
251 TCCAGAGGAG GCAGGCGAGT GAGCGAGTCC GAGGGGTGGC CGGGGCGAGT
301 GGTGGCGCGC CGAAGATGGT CGCCAAGCAA AGGATCCGTA TGGCCAACGA
351 GAAGCACAGC AAGAACATCA CCCAGCGCGG CAACGTCGCC AAGACCTCGA
401 GAAATGCCCC CGAAGAGAAG GCGTCTGTAG GACCCGTGTT ATTGGCTCTC
451 TTCATTTTGG TTGCTGTGG TTCTGCAATT TTCCAGATTA TTCAAAGTAT
501 CAGGATGGGC ATGTGAAGTG ACTGACCTTA AGATGTTTCC ATTCTCTCTG
551 GAATTTTAAC TTGAACCTCAT TCCTGATCTT TGATACCCGT GTTGAAAAACA
601 ATTCAGTAAA GCATCCTGCC TCAGAATGAC TTTCTATCA TGCTTCATGT
651 GTCATTCCAA GGTTCCTTCA TGAGTCATTC CAAGTTTCT AGTCCATACC
701 ACAGTGCCCT GCAAAAAACA CCACATGAAT AAAGCAATAA AATTTGATTG
751 TTAAGATACA GTAGTGGACC CTACTTATTC AGTCAATTAA GAGTAAGTTT
801 TTTTATGTGG TTATTAAAAC AGTATGAACA ATTAGTCTAA CTCTGCATAG
851 ACAGGTGCTA GATTTTGTTA ACCCAAATGT ATAAGTCAG TTAGCTTAAA
901 TTACAATTTG AAGTCTGTG GTTTTATAT AGCTAGGCAC TTTATTACTC
951 TTTTGAACGT AAAGCACACT CCTTATAGG TTCATGTAAC TGTCTGTAA
1001 TAAGGTGCTT ATAAATGGAA CAACTACACA GCCTAGTTT GCCACAACCT
1051 TTAGCATCTA AAAAGTTTAA AAAGCTTCTA AATGTCTAAT ATAAAGGGAG
1101 ATGCTTATAG CCACAACATC TATTTTACCA ATATTGTTTC CATTACACTA
1151 CCTTGGATTG TGCATGAGTG AGTATAGTAA CCCAAGATGC CATAAAAAAA
1201 AACTTGATCG TTTCTGACT TAATTAGTTA CTGTGGTTTC ACTAAAAGCT
1251 ACCGTGGTGG AGTGAAGTCA GTCAGGGAAG GTTTGTTTAT GTTACATTTA
1301 TTTCAACAGA ACTATTTTAA TATATCAAAG GGGTTTACTA TGCCAAACAA
1351 AATTCTAGGG AAAAATACTG CTAATAATGG ATGCCCTATC AGAACATGCT
1401 GTTGAGTCCA ATGTGCCATA AGACATTTTA GCATGTTAAA TAGCACTTTT
1451 AATAGCAAAA AAAGGCACAT CAACTGCGAA GTTATCCTTA GTTTGCAAAAT
1501 GCTTTTCTA GATTAATGAT TTTTCAATCA TTAGGGTACT AGACACATCA
1551 GCCTAAAGTG GCATCTGGAA TTGAATGGAT TTACTGATAA TGATCAGTCT
1601 TTAGTCTTCC CTTTGTATA TGACTTTATA GGTATGATT GATCAAATTT
1651 ACGTTTTACT AATGGTAAGG GTGAGGGTCA TAGGGCAGGT TTTGGGTTTT
1701 CTAGTACTGT TGAAAAGTGC AAGTATTGGC TATTTGTATA CTTAGCCATA
1751 ACTTGGTGAA AAAAACCTG AGCAGTGTCT ATGTATTAAT GCGTTGGAAA
1801 GAAAGCTGCT TGTGTTTGGT TTGTTAATTG CCTCAGGATA TTTCTTTTAA
1851 AATAAGCTGT TTTAAGAGGA ACAGAAGGGA AATCTGCTAC CTAGTCTATA
1901 CACAGCGTGA ACCTCACAGG GGGCTTCTGA TACCCTCAA CATGGAGAAC
1951 AGTAAGGGAG CAGAGTGGTT AAGGACTTTC AGGAACTTAA CTATTCTGGA
2001 ATAAGGAATG AATCAACTGA CCTTGGGCCA GCAGGTTTTT AACTAAATTG
2051 TTAATTCGCT TTCTACCCA GTTAATCAGT CTCTGTACTT GTTTCCTTTT
2101 GTAAACCAAG TGTCTTGGTT AACTAATTCT GTTTTATGGT TGTGCTAAAT
2151 TCATAGCAGG TGCCTTATTC TTTGCTTTTA GTCAAACCAT TCCATATCAG
2201 AATTTTCCTT GGTTTACTAT AGATATTTGG CTTTAAGTTG TTGTTTGTGT
2251 TTTTAAATGT ACAAATGTTCT GATAAATTG ACTGTTAAAT TGCTATAGCT
2301 AGCAATCATT TTACATATGT AAAAAATTGC ATTCCCTTTG TATTTTCATGT
2351 GTAAATTCACC AATTAAGTGC AGTTTATATT CAGGTTGGAT TATGCATGTT
2401 TAGGTAACG AAAGCTGTGT CTTACTTGAT TTATCTTTA AAAATAAAGT
2451 TCCCTGAATA TTTGAAAAAA AAAAAAATAA AAAAAAATAA
```

BLAST Results

Entry HSCDN13 from database EMBL:
H.sapiens (TL5) mRNA from LNCaP cell line
Score = 1075, P = 5.8e-41, identities = 219/221

Entry AF100470_1 from database TREMBLNEW:
gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds.
Score = 331, P = 3.9e-28, identities = 66/66, positives = 66/66, frame +1

Entry HSG19910 from database EMBL:
human STS A002B48.
Score = 530, P = 2.1e-17, identities = 108/109

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 316 bp to 513 bp; peptide length: 66
Category: strong similarity to known protein
Classification: Intracellular transport and traffic

1 MVAQQRIRMA NEKHSKNITQ RGNVAKTSRN APEEKASVGP WLLALFIFVV
51 CGSAIFQIIQ SIRMGM

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21116, frame 1

TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4, N = 1, Score = 331, P = 6.2e-30

TREMBL:AF100470_1 gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds., N = 1, Score = 331, P = 6.2e-30

>TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4
Length = 75

HSPs:

Score = 331 (49.7 bits), Expect = 6.2e-30, P = 6.2e-30
Identities = 66/66 (100%), Positives = 66/66 (100%)

Query: 1 MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ 60
MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ
Sbjct: 10 MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ 69

Query: 61 SIRMGM 66
SIRMGM

Sbjct: 70 SIRMGM 75

No Pedant data available

DKFZphtes3_21n23

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with strong similarity to rat 7acomp protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to rat 7acomp protein

on genomic level encoded by AF107885

Sequenced by LMU

Locus: /map="14q24.3"

Insert length: 3122 bp

Poly A stretch at pos. 3070, polyadenylation signal at pos. 3045

```
1  GGAAAACCTC GTGGGCTCAG CCCGGGACAA AGGGCCAGGG AAGTTGGGTG
51 GTTCTGTGCT TGGTCTGTCA ATGGAGGAGA TCAAAGTTT ACGAAGGGTG
101 AAGGAGGAGA ATGATCGGCG AGGTGGATTT ATTCGCATAT TTCCTACATC
151 TGAGACATGG GAAATATATG GGTCTACCT CGAGCATAAG ACCTCAATGA
201 ACTATATGCT GGCAACACGC CTCTCCAGG ACAGGGGAAA CCCAAGAAGA
251 AGTTATTGTA CAGGAAGAAC ACGAATGACT GCTGATGGAG CGCCAGAATT
301 GAAGATAGAG AGTCTGAATT CAAAGGCCAA GCTGCATGCT GCACTTTACG
351 AGAGGAAGCT CCTGTCTCTG GAGGTGCGAA AACGTAGACG ACGGAGTAGC
401 AGATTGAGGG CAATGAGGCC AAAATACCCA GTGATTACCC AACCAGCTGA
451 AATGAATGTT AAAACTGAGA CAGAGAGTGA AGAGGAGGAA GAAGTCGCAT
501 TAGATAATGA AGATGAAGAA CAGGAGGCTT CCCAGGAGGA GTCTGCAGGA
551 TTTCTTAGAG AAAATCAAGC CAAATATACA CCTCATTGA CAGCTTTGGT
601 AGAAAATACA CCCAAAGAAA ATTCCATGAA AGTTCGTGAA TGAATAATA
651 AAGGTGGACA CTGCTGCAAA CTTGAGACTC AGGAGCTAGA GCCTAAATTT
701 AACCTGATGC AGATTCTTCA AGATAATGGC AATCTTAGCA AAATGCAGGC
751 CCGAATAGCA TTCTCTGCCT ATCTCCAGCA TGTTCAAATT CGCCTGATGA
801 AAGACAGTGG CGGTCAAGC TTCAGTGCCA GTTGGGCTGC CAAAGAGGAT
851 GAACAGATGG AGCTGGTTGT TCGTTTCTCT AAGCGAGCAT CAAATAACCT
901 CCAGCATTC A CTGAGGATGG TATTACCCAG TCGACGATTG GCACTTCTGG
951 AACGCAGAAG AATCCTGGCC CACCAGCTGG GTGACTTTAT CATTGTATAC
1001 AACAAAGGAA CAGAACAAT GGCTGAAAAG AAATCAAAGA AGAAAGTTGA
1051 GGAAGAAGAG GAAGATGGGG TGAATATGGA AAACCTTCAG GAGTTCATCA
1101 GACAAAGCAAG TGAGGCTGAA CTGGAGGAGG TGTGACTTT TTATACCCAA
1151 AAGAACAAAT CTGCTAGTGT CTTCTGGGG ACTCACTCTA AAATTTCTAA
1201 GAACAACAAC AATTATTCTG ATAGTGGGGC AAAAGGTGAT CACCCTGAGA
1251 CTATAATGGA AGAAGTGAAA ATAAAGCCAC CTAACAGCA ACAGACGACA
1301 GAAATTCATT CTGATAAATT ATCTCGATTT ACCACTTCAG CAGAAAAAGA
1351 GGCAAAATTA GTTTATAGCA ATTCTCTCTC TGGTCTACT GCTACTCTGC
1401 AGAAAATTCC CAACACCCAT TTGTCTCTG TTAACAACCTC TGACCTCTCT
1451 CCAGGGCCTT GCCACCATTC TTTCTTATCT CAAATTCCTT CAGCTATCCC
1501 CAGCATGCCCT CACCAGCCAA CAATTTTACT GAACACAGTC TCTGCCAGTG
1551 CTCTCTCCCTG CCTACATCCC GGGGCACAGA ACATCCCAAG CCCTACTGGC
1601 CTGCCACGCT GTCGATCAGG AAGTCACACC ATTGGTCCCT TTTCTTCTCT
1651 CCAAAGTGCT GCACACATCT ATAGCCAGAA ACTGTCTCGT CCCTCTTCAG
1701 CAAAGGCAGG ATCGTGCTAT CTAACAAGC ATCATTGAGG AATAGCCAAA
1751 ACACAAAAG AGGGAGAAGA TGCTTCTTTA TATAGCAAA GGTACAACCA
1801 AAGTATGGTT ACAGCTGAAC TTCAGCGGCT AGCTGAGAAG CAGGCAGCGA
1851 GACAGTATTC TCCATCCAGC CACATCAACC TCCTCACCCA ACAGGTAACA
1901 AACCTGAATT TGGCAACTGG CATCATAAAC AGAAGCAGTG CTTAGCTCC
1951 CCCAACCCCTC CGACCCATCA TCAGTCTTAG TGGCCCGACA TGGTCTACAC
2001 AGTCAGACCC CCAAGCTCCC GAGAATCACT CCAGCTCTCC TGGAAGCAGG
2051 AGCCTGCAGA CAGGGGGATT TGCTGGGAA GGAGAAGTAG AAAACAACGT
2101 GTACAGCCAG GCTACAGGGG TGGTCCCCCA GCACAAGTAT CACCCACAG
2151 CAGGCAGCTA TCAGCTTCAA TTTGCCCTGC AGCAACTTGA ACAACAAAA
2201 CTTCACTCCC GGCAGCTCCT GGACCAGAGT CGAGCCCGGC ACCAGGCAAT
2251 CTTTGGCAGC CAGACACTAC CTAACCTCAA TTTATGGACA ATGAATAATG
2301 GTGCAGGTTG TAGAATTTC AGTGCCACAG CTAGTGGCCA GAAGCCAACC
2351 ACTCTGCCAC AAAAAGTGGT ACCACCTCCA AGTTCTTGCG CCTCCCTGGT
2401 TCCCAAACCC CCACCAACC ACGAACAAGT GCTCAGAAGG GCAACATCCC
2451 AGAAAGCTTC CAATACCCGC TTCAGATCCT CCTTTCAAAA CTATTTGTGG
2501 TATTTCTTCC AAGCAGTCAG CTGAAGTGA GACGACAGCC TACAACAAC
2551 TACATGCATC TGAAGTGTCT CTTGTAAATG AGCTTTTTTC AGAGCCAGAA
2601 TCATACTCTC CAGGAAATAT GGAGAAAGAA ACCTGAGGAG ATTGAAGTTT
2651 GCCAGGCACA AGGGCAAAAC TCAGACTGAA TGAATTGAA AGGGTGGGGC
2701 CAAAGATGTT GTAACCTGGG AGACTTCTCT GAAGAAAGAA AACTGTTTAA
```

```

2751 GAAACACAGA CTGAAGTCA GTACTTTTCC TTAATAGCT GAGATGACCT
2801 TCITTTACCCT GGGCTTAGGT GATTCTCATC AGGGTGACCT GAGTGGAAAGT
2851 TGGTGGTAAC GACTGTTCTG TGTCAGCACC CAGGACAGTG GTGTCTGTTA
2901 AGGCTGCCAG GGATTAGCAG GGAGGAAAGC CATCAGGACT GGGTAGCCTG
2951 GTAGCACCAA ATCCCAATTA ATGTTACCTG AACATGTGGT GAGGTCAGCC
3001 GTATGATGAA AGATGTTTAA GAGATTAATG TCAGAAGAAT ATGAAAATAA
3051 ACACCGGCTT AAAAAATGTT AAAAAAATAA AAAAAAATAA AAAAAAATAA
3101 AAAAAAATAA AAAAAAATAA AA

```

BLAST Results

Entry AF107885 from database EMBL:
Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes.
Score = 3042, P = 3.0e-219, identities = 610/612
5 exons matching 1893-3070

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 71 bp to 2521 bp; peptide length: 817
Category: strong similarity to known protein

```

1 MEEIKVLRV KEENDRRGGF IRIFPTSETW EIYGSYLEHK TSMNYMLATR
51 LFQDRGNPRR SLLTGRTRMT ADGAPELKIE SLNSKAKLHA ALYERKLLSL
101 EVRKRRRRSS RLRAMRPKYP VITQPAEMNV KTESEEEEE EVALDNEDEE
151 QEASQESAG FLRENQAKYT PSLTALVENT PKENSMKVRE WNNKGGHCCK
201 LETQELEPKF NLMQILQDNG NLSKMQARIA FSAYLQHVQI RLMKDSGGQT
251 FSASWAAKED EQMELVVRFL KRASNNLQHS LRMVLPSSRL ALLERRRILA
301 HQLGDFIIVY NKETEQMAEK KSKKKVEEEE EDGVMENFQ EFIRQASEAE
351 LEEVLTFYTO KNKSASVFLG THSKISKNNN NYSDSGAKGD HPETIMEEVK
401 IKPPKQQQTT EIHSKLSRF TTSAEKEAKL VYNSSSSGPT ATLQKIPNTH
451 LSSVTTSDLS PGPCHHSSLS QIPSAIPSMQ HQPTILLNTV SASASPCPLHP
501 GAQNIPSPPTG LPRCRSGSHT IGFSSSFQSA AHIYSQKLSR PSSAKAGSCY
551 LNKHHSGLAK TQKEGEDASL YSKRYNQSMV TAEQLRLAEK QAARQYSPSS
601 HINLLTQQVT NLNLATGIIN RSSASAPPTL RPIISPSGPT WSTQSDPQAP
651 ENHSSSPGSR SLQTGGFAWE GEVENNVYSQ ATGVVPQHKY HPTAGSYQLQ
701 FALQOLEQQK LQSRQLLDQS RARHQAFIFS QTLFNSNLWT MNNGAGCRIS
751 SATASGQKPT TLPQKVVPVP SSCASLVPKP PPNHEQVLRN ATSQKASNTR
801 FRSSFQNYLW YFFQAVS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2ln23, frame 2

TREMBL:AF064856_1 product: "7aomp protein"; Rattus sp. 7aomp protein mRNA, complete cds., N = 1, Score = 1845, P = 2.2e-190

TREMBL:AF107885_3 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 443, P = 5.3e-41

TREMBL:AF107885_4 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 265, P = 8.2e-22

>TREMBL:AF064856_1 product: "7aomp protein"; Rattus sp. 7aomp protein mRNA, complete cds.
Length = 436

HSPs:

Score = 1845 (276.8 bits), Expect = 2.2e-190, P = 2.2e-190
Identities = 369/435 (84%), Positives = 395/435 (90%)

Query:	115	MRPKYPVITQPAEMNVKTESEEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLT	174
		MRPKYPVIT PAEMN+KTETSEEEEEV LDNEDEEQEASQESAG L ENQAKYTPSLT	
Sbjct:	1	MRPKYPVITLPAEMNIKTETSEEEEEVLGDNEDEEQEASQESAGSLAENQAKYTPSLT	60
Query:	175	ALVENTPKENSMKVREWNNGGGHCKLETQLEPKFNLMLQILQDNGNLSKMQARIAFSAY	234
		+VEN+P+EN+MKV EW NKG CCK+ETQE E KFNLMQILQDNGNLSK+QAR+AFSAY	
Sbjct:	61	VIVENSPRENAMKVAEWTNKGESCCKITQETEPESKFNLMLQILQDNGNLSKVQARLAFSAY	120
Query:	235	LQHVQIRLMKDSGGQTFASWAAKEDEQOMELVVRFLKRASNQLQHSRLMVLPSRRLALLE	294
		LQHVQ+RL KDSGGQT S SWAAKEDEQOMELVVRFLKRAS+NLQHSRLMVLPSRRLALLE	
Sbjct:	121	LQHVQVRLTKDSGAEGQTLSPGAKEDEQOMELVVRFLKRASSNLQHSRLMVLPSRRLALLE	180
Query:	295	RRRILAHQLGDFIIVYNKETEOMAEEKSKKKVEEEEEDGVNMENFQEFIRQASAELEEV	354
		RRRILAHQLGDFI+VYNKETEOMAEEKSKKK+EEEEEDGVN E+FQEFIRQASAELEEV	
Sbjct:	181	RRRILAHQLGDFIIVVYNKETEOMAEEKSKKKLEEEEEDGVNAESFQEFIRQASAELEEV	240
Query:	355	LTFYTKNKSASVFLGTHSKISKNNNYSDSGAKGDHPETIMEEVKIKPKPKQQQTEIHS	414
		LTFYTKNKSASVFLGTHSK SKN+++YSDSGAKGDHPETI +EVKIK PKQQQ TEIHS	
Sbjct:	241	LTFYTKNKSASVFLGTHSKSKNSSSYSDSGAKGDHPETI-QEVKIKPKPKQQQATEIHS	299
Query:	415	DKLSRFTTSAEKEAKLVYSNSS--GPTATL-QKIPNTHLSSV-TTSDLSPGCHHSSLS	470
		DKLSRFTTSA KEAKLVY+N SS GP A L Q++P+THLSS+ TTS LS GP HHSSLS	
Sbjct:	300	DKLSRFTTSAGKEAKLVYNTCCSFSFGPAVVLQRLPSTHLSSIITTLSSSGPGHHSSLS	359
Query:	471	QIPSAIPSMHPQPTILLNTVSASPLCHPGAQNIPSPTGLPRCRSGSHTIGPFSSFQSA	530
		QI AIPSMHPQ +LLN V SASP +HPG N+ SP GLPRCRSGS+TIGPFSSFQSA	
Sbjct:	360	QIPSAIPSMHPQSALLNPVPDSASPPVHPGTENV-SPAGLPRCRSGSYTIGPFSSFQSA	418
Query:	531	AHIIYSQKLSRPSSAKAG	547
		AHIIYSQKLSRPSSAKAG	
Sbjct:	419	AHIIYSQKLSRPSSAKAG	435

Pedant information for DKFZphtes3_21n23, frame 2

Report for DKFZphtes3_21n23.2

```
[LENGTH]          817
[MW]               91522.09
[pI]               9.32
[HOMOL]            TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA,
complete cds. 1e-166
[PROSITE]          MYRISTYL          6
[PROSITE]          CAMP_PHOSPHO_SITE 4
[PROSITE]          CK2_PHOSPHO_SITE  12
[PROSITE]          TYR_PHOSPHO_SITE   1
[PROSITE]          PKC_PHOSPHO_SITE   15
[PROSITE]          ASN_GLYCOSYLATION  7
[KW]               Alpha_Beta
[KW]               LOW_COMPLEXITY      13.83 %
```

[illegible]

```

SEQ      TTSAEKEAKLVYNSSSSGPTATLQKIPNTHLSSVTTSDLSPGPCHHSSLQIPSAIPSPM
SEG      .....
PRD      hhhhhhhheeeccccccccceeecccccccccccccccccccccccccccccccccccc

SEQ      HQPTILLNTVSASAPCLHPGAQNIPTGLPRCRSGSHTIGPFSSFQSAAHYISQKLSR
SEG      .....
PRD      cccceeeccccccccccccccccccccccccccccccccccccccccchhhhhhhhhccc

SEQ      PSSAKAGSCYLNKHHSGIAKTQKEGEDASLYSKRYNQSMVTAELQRLAEKQARQYSPSS
SEG      .....
PRD      cccccceeeccccccccccccccccceeeccchhhhhhhhhhhhhhhhhhhhhccc

SEQ      HINLLTQQVTNLNLATGIINRSSASAPPTLRPIISPSGPTWSTQSDPQAPENHSSSPGSR
SEG      ..xxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccceeecccccccccccccccccccccccccccc

SEQ      SLQTGGFAWEGEVENNVYSQATGVVPQHKYHPTAGSYQLQALQOLEQOKLQSRQLDQS
SEG      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccceeeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      RARHQAIFGSQTLPSNLWTMNGAGCRISSATASGQKPTTLPQKVVPSPSSCASLVKPK
SEG      .....
PRD      hhhhhhhhhccccccccceeeccccceeeccccccccceeeccccceeecccc

SEQ      PPNHEQVLRRATSQKASNTRFRSSFQNYLWYFFQAVS
SEG      .....
PRD      cccccchhhhhhhhhccccccccccccceeecccc

```

Prosites for DKFZphtes3_21n23.2

PS00001	221->225	ASN_GLYCOSYLATION	PDOC00001
PS00001	362->366	ASN_GLYCOSYLATION	PDOC00001
PS00001	381->385	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00001	620->624	ASN_GLYCOSYLATION	PDOC00001
PS00001	652->656	ASN_GLYCOSYLATION	PDOC00001
PS00004	106->110	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	271->275	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	789->793	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	109->112	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	185->188	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	543->546	PKC_PHOSPHO_SITE	PDOC00005
PS00005	561->564	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00005	629->632	PKC_PHOSPHO_SITE	PDOC00005
PS00005	793->796	PKC_PHOSPHO_SITE	PDOC00005
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	347->351	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	422->426	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	561->565	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00007	563->572	TYR_PHOSPHO_SITE	PDOC00007
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	248->254	MYRISTYL	PDOC00008
PS00008	510->516	MYRISTYL	PDOC00008
PS00008	557->563	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	756->762	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_21n23.2)

DKFZphtes3_22c23

group: testes derived

DKFZphtes3_22c23 encodes a novel 223 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, 3 EST hits (two from a testis library)

Sequenced by LMU

Locus: /map="9q34"

Insert length: 1113 bp

Poly A stretch at pos. 1073, polyadenylation signal at pos. 1055

```
1 GGTGGGCAAA GGCATCTTCC TCTGGGAAGG ACTGGCACAA GCACCTGGTC
51 CTTGGGTTGT GTGCCTGGGA GGCCGGGATC AGGGCTGGCC CTCTTTCTCC
101 CTGGCAAAGC AAAACCTCCC TTTTACTACT ATCAAGGGGA AGTAACTTGA
151 AGGTGCCTGT GGCAGGCAGC ACCTTGAGCC AACAGGAACC ATTGACATGC
201 GAGGCCCAGG GCAGGCAGAC TGTGCAGTGG CCATTGGGCG GCCCCTCGGG
251 GAGGTGGTGA CCCTCCGCGT CTTGAGAGT TCTCTCAACT GCAGTGCGGG
301 GGACATGTTG CTGCTTTGGG GCCGGCTCAC CTGGAGGAAG ATGTGCAGGA
351 AGCTGTTGGA CATGACTTTC AGCTCCAAGA CCAACACGCT GGTGGTGAGG
401 CAGCGCTGCG GCGGCCAGG AGTGGGGTG CTGCTGCGGT ATGGGAGCCA
451 GCTTGCTCCT GAAACCTTCT ACAGAGAATG TGACATGCAG CTCTTTGGGC
501 CTTGGGGTGA AATCGTGAGC CCCTCGCTGA GTCCAGCCAC GAGTAATGCA
551 GGGGGCTGCC GGCTCTTCAT TAATGTGGCT CCGCACGCAC GGATTGCCAT
601 CCATGCCCTG GCCACCAACA TGGGCGCTGG GACCGAGGGA GCCAATGCCA
651 GCTACATCTT GATCCGGGAC ACCCACAGCT TGAGGACCAC AGCGTTCCAT
701 GGGCAGCAGG TGCTCTACTG GGAGTCAGAG AGCAGCCAGG CTGAGATGGA
751 GTTCAGCAGG GGCTTCTCTA AGGCTCAGGC CAGCCTGCGG GGCCAGTACT
801 GGACCCTCCA ATCATGGGTA CCGGAGATGC AGGACCCTCA GTCCTGGAAG
851 GAAAAGGAAG GAACCTGAGG GTCATTGAAC ATTTGTTCCG TGCTTGCCA
901 GCCCTGGAGG GTTGACCCCT GGTCTCAGTG CTTTCCAATT CGAAGTTTTT
951 CCAATCTTAG GTATCTACTT TAGAGTCTTC TCCAATGTCC AAAAGGCTAG
1001 GGGGTTGGAG GTGGGGACTC TGGAAAAGCA GCCCCCATTT CCTCGGGTAC
1051 CAATAAATAA AACATGCAGG CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA
1101 AAAAAAAAAA AAA
```

BLAST Results

Entry HSAC1644 from database EMBL:
Genomic sequence from Human 9q34, complete sequence.
Score = 2072, P = 8.8e-225, identities = 422/430
5 exons Bp 41969-38232

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 197 bp to 865 bp; peptide length: 223
Category: putative protein

```
1 MRGPGQADCA VAIGRPLGEV VTLRVLESSL NCSAGDMLLL WGRLTWRKMC
51 RKLLDMTFSS KTNLTVVRQR CGRPGGGVLL RYGSQAPET FYRECDMQLF
101 GPWGEIVSPS LSPATSNAGG CRLFINVAPH ARIAIHALAT NMGAGTEGAN
151 ASYILIRDTL SLRTTAFHQO QVLYWESESS QAEMEFSEGF LKAQASLRGQ
201 YWTLQSWVPE MQDPQSWKKG EGT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22c23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_22c23, frame 2

Report for DKFZphtes3_22c23.2

```

[LENGTH]      223
[MW]           24546.19
[pI]           8.57
[PROSITE]      MYRISTYL      4
[PROSITE]      CK2_PHOSPHO_SITE      2
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           Alpha_Beta

SEQ    MRGPGQADCAVAIGRPLGEVVTLRVLESSLNCSAGDMLLLWGRLTWRKMCRLDMTFSS
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    KTNTLVVRQRCGRPGGVLLRYGSQLAPETFYRECDMQLFGPWGEIVSPSLSPATSNAGG
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    CRLFINVAPHARIAIHALATNMGAGTEGANASYILIRDTHSLRTTAFHGQQVLYWESESS
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    QAEMEFSEGFLKAQASLRGQYWTLSWVPEMQDPQSWKGKEGT
PRD    hhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Prosites for DKFZphtes3_22c23.2

```

PS00001      31->35    ASN_GLYCOSYLATION      PDOC00001
PS00001      150->154  ASN_GLYCOSYLATION      PDOC00001
PS00005      22->25    PKC_PHOSPHO_SITE      PDOC00005
PS00005      45->48    PKC_PHOSPHO_SITE      PDOC00005
PS00005      59->62    PKC_PHOSPHO_SITE      PDOC00005
PS00005      161->164  PKC_PHOSPHO_SITE      PDOC00005
PS00005      196->199  PKC_PHOSPHO_SITE      PDOC00005
PS00005      216->219  PKC_PHOSPHO_SITE      PDOC00005
PS00006      33->37    CK2_PHOSPHO_SITE      PDOC00006
PS00006      180->184  CK2_PHOSPHO_SITE      PDOC00006
PS00008      5->11     MYRISTYL              PDOC00008
PS00008      145->151  MYRISTYL              PDOC00008
PS00008      148->154  MYRISTYL              PDOC00008
PS00008      199->205  MYRISTYL              PDOC00008

```

(No Pfam data available for DKFZphtes3_22c23.2)

DKFZphtes3_22g2

group: nucleic acid management

DKFZphtes3_22g2 encodes a novel 1230 amino acid protein with nearly identical to rat TIP120.

TATA-binding protein TBP is a central component for transcriptional regulation and is a target for various transcription regulators. TBP-interacting protein 120 (TIP120) is a protein interacting with the TATA-binding protein (TBP). The novel protein is the human ortholog of rat TIP120. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP.

The new protein can find application in modulation of gene transcription.

KIAA0829, complete cds, nearly identical to rat TIP120

complete cDNA, complete cds, EST hits,

Sequenced by LMU

Locus: /map="387.3 cR from top of Chr12 linkage group"

Insert length: 5387 bp

Poly A stretch at pos. 5352, polyadenylation signal at pos. 5335

```
1 GGGAGCGAGT GCGGAGCGAG TGGGAGCGAG ACGGCCCTGA GTGGAAGTGT
51 CTGGCTCCCC GTAGAGGCCCT TTCTGTACGC CCCGCCGCC ATGAGCTCGT
101 TCTCAGCGCA ACAGCGCCCT CGTTAGGCTG GCTCTGTAGC CTCGGCTTAC
151 CCCGGGACAG GCCACGCCT CGCCAGGGAG GGGGCAGCCC GTCGAGGCGC
201 CTCCCTAGTC AGCGTCGGCG TCGCGCTGCG ACCCTGGAAG CGGGAGCCGC
251 CGCGAGCGAG AGGAGGAGCT CCAGTGGCGG CGGCGGCGGC GGCAGCGGCA
301 GCGGGCAGAG GCTCCAGCAG CGCCAGCAGG CGGGATCGAG GCCGTCAACA
351 TGGGAGCGCG CTCGTACCAC ATTTCCAATT TGCTGGAAAA AATGACATCC
401 AGCCGACAAG ACTTTAGGTT TATGGCTACA AATGATTGA TGACGGAAC
451 GCAGAAAGAT TCCATCAAGT TGGATGATGA TAGTGAAAGG AAAGTAGTGA
501 AAATGATTTT GAAGTTATTG GAAGATAAAA ATGGAGAGGT ACAGAAATTTA
551 GCTGTCAAAT GTCTTGGTCC TTTAGTGAGT AAAGTGAAAG AATACCAAGT
601 AGAGACAATT GTAGATACCC TCTGCACTAA CATGCTTTCT GATAAAGAAC
651 AACTTCGAGA CATTTCAGT ATTGGTCTTA AAACAGTAAT TGGAGAAC
701 CCTCCAGCTT CCAGTGGCTC TGCATTAGCT GCTAATGTAT GTAAAAAGAT
751 TACTGGACGT CTTACAAGTG CAATAGCAAA ACAGGAAGAT GTCTCTGTTC
801 AGCTAGAAGC CTTGGATATT ATGGCTGATA TGTTGAGCAG GCAACGAGGA
851 CTTCTTGTTA ATTTCCATCC TTCAATTCTG ACCTGTCTAC TTCCCCAGTT
901 GACCAGCCCT AGACTTGCAG TGAGGAAAAG AACCATTATC GCTCTGGGCC
951 ATCTGGTTAT GAGCTGTGGA AATATAGTTT TTGTAGATCT TATTGAACAT
1001 CTGTTGTCAG AGTTGTCCAA AAATGATTCT ATGTCAACAA CAAGAACCTA
1051 CATACAATGT ATTGCTGCTA TTAGTAGGCA AGCTGGTCAT AGAATAGGTG
1101 AATACCTTGA GAAGATAATT CCTTTGGTGG TAAAAATTTG CAATGTAGAT
1151 GATGATGAAT TAAGAGAGTA CTGTATTCAA GCCTTTGAAT CATTGTGAAG
1201 AAGATGTCCT AAGGAAGTAT ATCCTCATGT TTCTACCATT ATAAATATTT
1251 GTCTTAAATA TCTTACCTAT GATCCAAATT ATAATTACGA TGATGAAAGT
1301 GAAGATGAAA ATGCAATGGA TGCTGATGGT GGTGATGATG ATGATCAAGG
1351 GAGTGATGAT GAATACAGTG ATGATGATGA CATGAGTTGG AAAGTGAGAC
1401 GTGCAGCTGC GAAGTGCTTG GATGCTGTAG TTAGCACAAG GCATGAAATG
1451 CTTCCAGAAT TCTACAAGAC CGTCTCTCCT GCACTAATAT CCAGATTTAA
1501 AGAGCGTGAA GAGAATGTAA AGGCAGATGT TTTTCACGCA TACCTTTCTC
1551 TTTTGAAGCA AACTCGTCTT GTACAAAAGT GGCTATGTGA CCCTGATGCA
1601 ATGGAGCAGG GAGAAACACC TTAACAATG CTTCAGAGTC AGGTTCCCAA
1651 CATTGTTAAA GCTCTTCACA AACAGATGAA AGAAAAAGT GTGAAGACCC
1701 GACAGTGTTG TTTTAACATG TTAAGTGAGC TGGTAAATGT ATTACCTGGG
1751 GCCCTAACTC AACACATTCC TGTACTTGTA CCAGGAATCA TTTTCTCACT
1801 GAATGATAAA TCAAGCTCAT CGAATTTGAA GATCGATGCT TTGTCATGTC
1851 TATACGTAAT CCTCTGTAAC CATTCTCCTC AAGTCTTCCA TCCTCACGTT
1901 CAGGCTTTGG TTCTCCAGT GGTGGCTTGT GTTGAGAGCC CATTTTACAA
1951 AATTACATCT GAAGCACTTC TTGTTACTCA ACAGCTTGTC AAAGTAATTC
2001 GTCCTTTAGA TCAGCCTTCC TCGTTTGATG CAACTCCTTA TATCAAAGAT
2051 CTATTTTACCT GTACCATTAA GAGATTAAAA GCAGCTGACA TTGATCAGGA
2101 AGTCAAGGAA AGGGCTATTT CCTGTATGGG ACAAAATTAT TGCAACCTTG
2151 GAGACAATTT GGGTTCTGAC TTGCCTAATA CACTTCAGAT TTTCTTGGAG
2201 AGAGCTAAAG ATGAAATTAC CAGGTAACT ACAGTAAAGG CATTGACACT
2251 GATTGCTGGG TCACCTTTGA AGATAGATTG GAGGCCTGTT CTGGGAGAA
2301 GGGTTCTTAT CTTGCTTCA TTTCTTAGAA AAAACCAGAG AGCTTTGAAA
2351 CTGGGTACTC TTTCTGCCCT TGATATTCTA ATAAAAAAT ATAGTGACAG
2401 CTTGACAGCT GCCATGATTG ATGCAGTTCT AGATGAGCTC CCACCTCTTA
2451 TCAGCGAAAG TGATATGCAT GTTTCACAAA TGGCCATCAG TTTTCTTACC
2501 ACTTTGGCAA AAGTATATCC CTCCTCCCTT TCAAAGATAA GTGGATCCAT
2551 TCTCAATGAA CTTATTGGAC TTGTGAGATC ACCCTTATTG CAGGGGGGAG
```

```

2601 CTCTTAGTGC CATGCTAGAC TTTTCCAAG CTCTGGTTGT CACTGGAACA
2651 AATAATTTAG GATACATGGA TTTGTTGCGC ATGCTGACTG GTCCAGTTTA
2701 CTCTCAGAGC ACAGCTCTTA CTCATAAGCA GTCTTATTAT TCCATTGCCA
2751 AATGTGTAGC TGCCCTTACT CGAGCATGCC CTAAAGAGGG ACCAGCTGTA
2801 GTAGGTCAGT TTATTCAAGA TGTCAAGAAC TCAAGGTCTA CAGATTCCAT
2851 TCGTCTCTTA GCTCTACTTT CTCTTGGAAG AGTTGGGCAT CATATTGACT
2901 TAAGTGGACA GTTGGAACATA AAATCTGTAA TACTAGAAGC TTTCTCATCT
2951 CCTAGTGAAG AAGTCAAAATC AGCTGCATCC TATGCATTAG GCAGCATTAG
3001 TGTGGGCAAC CTTCTGAAT ATCTGCCGTT TGTCTGCAA GAAATAACTA
3051 GTCAACCCAA AAGGCAGTAT CTTTACTTC ATTCTTGAA GGAATTATT
3101 AGCTCTGCAT CAGTGGTGGG CCTTAAACCA TATGTTGAAA ACATCTGGGC
3151 CTTATTACTA AAGCACTGTG AGTGTGCAGA GGAAGGAACC AGAAATGTTG
3201 TTGCTGAATG TCTAGGAAAA CTCACCTCTAA TTGATCCAGA AACTCTCCTT
3251 CCACGGCTTA AGGGGTACTT GATATCAGGC TCATCATATG CCCGAAGCTC
3301 ACTGGTTACG GCTGTGAAAT TTACAATTTC TGACCATCCA CAACCTATTG
3351 ATCCACTGTT AAAGAAGTGC ATAGGTGATT TCCTAAAAAC TTTGGAAGAC
3401 CCAGATTTGA ATGTGAGAAG AGTAGCCTTG GTCACATTTA ATTCAGCAGC
3451 ACATAACAAG CCATCATTAA TAAGGGATCT ATTGGATACT GTTCTTCCAC
3501 ATCTTTACAA TGAACAAAAA GTTAGAAAGG AGCTTATAAG AGAGGTAGAA
3551 ATGGGTCCAT TTAACATAC GGTGATGAT GGTCTGGATA TTAGAAAGGC
3601 AGCATTGAG TGTATGTACA CACTTCTAGA CAGTTGTCTT GATAGACTTG
3651 ATATCTTTGA ATTTCTAAAT CATGTTGAAG ATGTTTGAA GGACCATTAT
3701 GATATTAAGA TGCTGACATT TTTAATGTTG GTGAGACTGT CTACCCTTTG
3751 TCCAAGTGCA GACTGCAGA GGTGGACCG ACTTGTGAG CCATTACGTG
3801 CAACATGTAC AACTAAGGTA AAGGCAAACT CAGTAAAGCA GGAGTTTGAA
3851 AAACAAGATG AATTAAGCG ATCTGCCATG AGAGCAGTAG CAGCACTGCT
3901 AACCATTCCA GAAGCAGAGA AGAGTCCACT GATGAGTGAA TTCCAGTCAC
3951 ACATCAGTTC TAACCTGAG CTGGCGGCTA TCTTTGAAAG TATCCAGAAA
4001 GATTTCATCAT CTAATAACTT GGAATCAATG GACACTAGTT AGATGTTTGT
4051 TCACCATGGG GACCATTACA TATGACCATA CAATGCACTG AATTGACAGG
4101 TTAATCATAA GACATGGAAA GAGAAGTGTC TAAAGCTTC AAAATGTTCC
4151 ACTTTTTTTT CCTTCATGGA GACTGTTTGT TTGGCTTTCT TCCATTGTTG
4201 TTTTGTAGC ATTTATTTC AATATGTTG TTTCCATAAT CCAGAGGTTG
4251 TAAACCACT AGTGTTTTAG TGGTTACAGC AACATTTGAA ATGGAAACTA
4301 AAAGTTAGGA TTTTATGGAG TATGGAGATA GGGTCCAGTA TCTATTTACC
4351 CTGTAATGTT TAGGATTAAA ATGTTAAAA TTTGTACCA TGAATTTCTT
4401 TCTTTTATAA ATTTCTCAT TTAATAATCA AAAATCTTGC AAAACAAAAA
4451 CCATGTTTCT TTTTCTTGT TAACTTTTG TTTTCAGCA CATAAATTGA
4501 TTTTATAGCT GCAGACAAGA ATATCCATAT AAGATTGTTT AACCATTTCA
4551 GAGAGTTTGG CAATTTTAA AAGATAATAA GGTATCATT TTAAGTATGA
4601 AAATTAAACA TATCCCTGTT GCGCACACTA ATTTGCAATG AGTAAGTTTA
4651 CAAATATGTA TCGTCTGTAA AGCAGCATGT GCAGATTATT CATAATATAG
4701 AAGTTAAAAA AAGTATTAGT GCAATTTTCA GATATTATT TTTGCACAGA
4751 AAACACATTA TCTGGAGAGA AAGAAAGGAG AATTTTGTAG ACTTGGGTTT
4801 TCTTAATGCC AGTGTGAATT TGCAGATGTT TTCAGAAAAT CAAGTCACAG
4851 TAACAATTTG CCACTTTTTT CTATTATAAA TCTTCTACT TAAATTTTGA
4901 ATATTTAGTT TTTCTCAGTT ACCCATTTGT GTGTGTGTA TTCCACTTAG
4951 AAATCTTAA AACCAAGATT TTCTTTCATT CCGTTTGGAT GTCTACATTC
5001 CTTATCAAAG GATATAAATA CTGTGTATGC TTTTGAATTT TATTTTATAG
5051 AAAATTCTGA AGCCAGCTAT CACAGGTTTG TTAGCTAATA ATAGTATTTT
5101 CTTTATGTTG AGTTAGGTTT TTCCCATCT CCGTGTAGAGC GAATTTACAT
5151 ATTGTATGG GTAAGTGTTC ACTACTTTTC CTGATTAGG GATCTGTGCT
5201 GGGGAACAAA GCTTTTGCAG TACCTTATAT TGTAGTTAAA ATTTATTTTA
5251 ACATATCCTT CAGTGAGCTC ATTTACACT GTAGCCTCTT CCTTAAATTT
5301 TGTGGTGCTC CTGTAACAGT AAGAACTAAT TCTGAAATAA AAGACATCTC
5351 CTAACAAAAA AAAAAA AAAAAA AAAAAA

```

BLAST Results

Entry HS793345 from database EMBL:

human STS WI-12457.

Score = 1985, P = 1.3e-83, identities = 433/460

Medline entries

97127450:

Molecular cloning of a novel 120-kDa TBP-interacting protein.

Peptide information for frame 2

ORF from 350 bp to 4039 bp; peptide length: 1230

Category: known protein
 Classification: Nucleic acid management

```

1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV
51 KMILKLEDEK NGEVQNLAVK CLGPLVSKVK EYQVETIVDT LCTNMLSDEK
101 QLRDISSIGL KTVIGELPPA SSGSALAANV CKKITGRITS AIAKQEDVSV
151 QLEALDIMAD MLRSQGGLLV NFHPSILTCL LPQLTSPRLA VRKRTIIALG
201 HLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA ISRQAGHRIG
251 EYLEKIIPLV VKFCNVDDDE LREYCIQAFE SFVRRCPKEV YPHVSTIINI
301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDD DDQGSDEYS DDDMSWKVR
351 RAAAKCLDAV VSTRHEMLPE FYKTVSPALI SRFKEREENV KADVHFAYLS
401 LLKQTRPVQS WLCDDPDAMEQ GETPLTMLQS QVPNIVKALH KQMKESVKT
451 RQCCFNMLTE LVNVLPALT QHIPVLVPGI IFSLNDKSSS SNLKIDALSC
501 LYVILCNHSP QVFPVHVQAL VPPVAVCGD PFYKITSEAL LVTQQLVKVI
551 RPLDQSSFD ATPYIKDLFT CTIKRLKAAD IDQEVKERAI SCMQIICNL
601 GDNLGSDFPN TLQIFLERLK NEITRLTTVK ALTLIAGSPL KIDLRPVLGE
651 GVPIIASFLR KNQRAKLGK LSALDILIKN YSDSLTAMI DAVLDELPLP
701 ISESDMHVSQ MAISFLTTLA KVPSSLSKI SGSILNELIG LVRSPLLQGG
751 ALSAMLDFFQ ALVVTGTNNL GYMDLLRMLT GPVYSQSTAL THKQSYYSIA
801 KCVAALTRAC PKEGPAVVGQ FIQDVKNRSR TDSIRLLALL SLGEVGHHD
851 LSGQLELKSIV ILEAFSSPSE EVKSAASYAL GSISVGNLPE YLPFVLQEIT
901 SQPKROYLLL HSLKEIISA SVVGLKPYVE NIWALLKHC ECAEEGTRNV
951 VAECGLKLT IDPETLLPRL KGYLISGSSY ARSSVVTAVK FTISDHPQPI
1001 DPLLKNCIGD FLKLTLEDPL NVRRLVLT NSAAHNKPSL IRDLDTVLV
1051 HLYNETKVRK ELIREVEMGP FKHTVDDGLD IRKAAFECEY TLDSCDLRL
1101 DIFEFLNHVE DGLKDHYDIK MLTFLMLVRL STLCPSAVLQ RLDRLVEPLR
1151 ATCTTKVKAN SVKQFEKQD ELKRSAMRAV AALLTIPEAE KSPLMSEFQS
1201 QISSNPELAA IFESIQKDS STNLESMDTS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22g2, frame 2

TREMBL:AB020636_1 gene: "KIAA0829"; product: "KIAA0829 protein"; Homo sapiens mRNA for KIAA0829 protein, partial cds., N = 1, Score = 5986, P = 0

TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds., N = 1, Score = 6203, P = 0

>TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds.
 Length = 1,230

HSPs:

Score = 6203 (930.7 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 1227/1230 (99%), Positives = 1228/1230 (99%)

```

Query: 1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60
      1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60
Sbjct: 1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60

Query: 61 NGEVQNLAVK CLGPLVSKVKEY QVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120
      61 NGEVQNLAVK CLGPLVSKVKEY QVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120
Sbjct: 61 NGEVQNLAVK CLGPLVSKVKEY QVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120

Query: 121 SSGSALAANV CKKITGRITSAIAKQEDVSVQLEALDIMAD MLRSQGGLLV NFHPSILTCL 180
      121 SSGSALAANV CKKITGRITSAIAKQEDVSVQLEALDIMAD MLRSQGGLLV NFHPSILTCL 180
Sbjct: 121 SSGSALAANV CKKITGRITSAIAKQEDVSVQLEALDIMAD MLRSQGGLLV NFHPSILTCL 180

Query: 181 LPQLTSPRLAVRKRRTI IALGHLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA 240
      181 LPQLTSPRLAVRKRRTI IALGHLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA 240
Sbjct: 181 LPQLTSPRLAVRKRRTI IALGHLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA 240

Query: 241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300
      241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300
Sbjct: 241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300

Query: 301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDDQGSDEYS DDDMSWKV RRAAAKCLDAV 360
      301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDDQGSDEYS DDDMSWKV RRAAAKCLDAV 360
Sbjct: 301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDDQGSDEYS DDDMSWKV RRAAAKCLDAV 360

Query: 361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWL CDDPDAMEQ 420
      361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWL CDDPDAMEQ 420
Sbjct: 361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWL CDDPDAMEQ 420

```

Query: 421 GETPLTMLQSQVNPVIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI 480
 GETPLTMLQSQVNPVIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI
 Sbjct: 421 GETPLTMLQSQVNPVIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI 480

Query: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL 540
 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL
 Sbjct: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL 540

Query: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600
 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
 Sbjct: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600

Query: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660
 GDNLG DL NTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR
 Sbjct: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660

Query: 661 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPILISEDMHVSQMAISFLTTLA 720
 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPILISEDMHVSQMAISFLTTLA
 Sbjct: 661 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPILISEDMHVSQMAISFLTTLA 720

Query: 721 KVPYSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780
 KVPYSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT
 Sbjct: 721 KVPYSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780

Query: 781 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840
 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL
 Sbjct: 781 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840

Query: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900
 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
 Sbjct: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900

Query: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960
 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT
 Sbjct: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960

Query: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD 1020
 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD
 Sbjct: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD 1020

Query: 1021 NVRRVALVTFNSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080
 NVRRVALVTFNSAAHNKPSLIRDLLD+VLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
 Sbjct: 1021 NVRRVALVTFNSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080

Query: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140
 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
 Sbjct: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140

Query: 1141 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPMLSEFQS 1200
 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPMLSEFQS
 Sbjct: 1141 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPMLSEFQS 1200

Query: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230
 QISSNPELAAIFESIQRDSSSTNLESMDTS
 Sbjct: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230

Pedant information for DKFZphtes3_22g2, frame 2

Report for DKFZphtes3_22g2.2

[LENGTH] 1230
 [MW] 136376.58
 [pI] 5.52
 [HOMOL] TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for
 TIP120, complete cds. 0.0
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 5.28 %

SEQ MASASYHISNLEKMTSSDKDFRFRMATNDLMTLQKDSIKLDDDSERKVVKMILKLEDEK
 SEG
 PRD cccccchhhhhhhhhccccceeeehhhhhhhhhccccccccchhhhhhhhhhhhhcc
 MEM
 SEQ NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTIVIGELPPA
 SEGxxxx
 PRD cccccceeeeeeceeeehhhhhhhhhhhccccchhhhhccccccccchhhhhhhhhcccc

MEM
SEQ SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL
SEG xxxxxxxx
PRD cccccchhhhhccchhhhhcccccchhhhhhhhhhhhhhhccceeeccchhhhh
MEM
SEQ LPQLTSPRLAVKRRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDMSMTTRTYIQCIAA
SEG
PRD hccccchhhhhhhhhhhheeecccccceehhhhhhhhhhhcccccchhhhhhhhhhh
MEM MMMMMMMMMMMMMM
SEQ ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPEVYPHVSTIINI
SEG
PRD hhhccccccccchhhhhhhheeeccchhhhhhhhhhhhhcccccceecchhhhh
MEM
SEQ CLKYLTYPNINYDDEDEDENAMADAGDDDDQGSDDDEYSDDDMSWKVRAAAKCLDAV
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh
MEM
SEQ VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHAYLSLLKQTRPVQSWLCDPDAMEQ
SEG
PRD hhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeccccc
MEM
SEQ GETPLTMLQSQVPNIVKALHKQKESVKTRQCCFNMLTELNVNLPGALTQHIPVLVPGI
SEG
PRD cccccchhhcccccceeeccce
MEM
SEQ IFSLNDKSSSNLKDALSCLYVILCNHSPQVFHPHVQALVPPVAVCGDPFYKITSEAL
SEG xxxxxxxxxxxxxxxx
PRD eeeccccccccchhhhhhhheeeccccccccceeeccceeecccccchhhhhhh
MEM
SEQ LVTQQLVKVIRPLDQSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
SEG
PRD hhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhheeecc
MEM
SEQ GDNLGSDLPNTLQIFLERLKNEITRLTVKALTLAGSPLKIDLRPVLGEGVPIIASFLR
SEG
PRD cccccccccchhhhhhhhhccchhhhhhhhhhhheeeccccccccceehhhhhhhhhhh
MEM
SEQ KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPISDMHVSQMAISFLTTLA
SEG
PRD hhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhcccccchhhhhhhhhhhhhhh
MEM
SEQ KVYPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT
SEG
PRD cccccceecchhhhhhhhhhhcccccchhhhhhhhhhhheeeccccchhhhhhhhhc
MEM
SEQ GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGGQFIQDVKNRSTDSIRLLALL
SEG
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhcccccchhhhhhh
MEM
SEQ SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
SEG
PRD hccccccccccccccccceeeccccchhhhhhhhhhhcccccchhhhhhhhhhh
MEM
SEQ SQPKRQYLLLSLKETISSASVVLKPYVENIWALLKHCECAEEGTRNVVAECLGKLT
SEG
PRD cccchhhhhhhhhhhhhcccccceehhhhhhhhhhhhhhhcccccceeecccccccc
MEM
SEQ IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQIDPLLKNCIGDFLKTLEDPD
SEG
PRD cccccccccccccccccchhhhhhhhhhhcccccchhhhhhhhhhhccccc
MEM
SEQ NVRRVALVTFNSAAHNKPSLIRDLLDTVLPPLHYNETKVRKELIREVEMGPFKHTVDDGLD
SEG
PRD cceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccch
MEM

```
SEQ  IRKAAFECCMYTLLDSCLDRLDIFEFLNHVEDGLKDHDIKMLTFLMLVRLSTLCPSAVLQ
SEG  .....
PRD  hhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RLDRLVEPLRATCTTKVKANSVKQEFQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
SEG  .....
PRD  hhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  QISSNPELAAIFESIQKSSSTNLESMDTS
SEG  .....
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....
```

(No Prosite data available for DKFZphtes3_22g2.2)

(No Pfam data available for DKFZphtes3_22g2.2)

DKFZphtes3_22n13

group: testes derived

DKFZphtes3_22n13 encodes a novel 677 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

dJ1042K10.3, complete

Sequenced by LMU

Locus: /map="22q13.1-13.2"

Insert length: 3353 bp

Poly A stretch at pos. 3315, polyadenylation signal at pos. 3298

```
1 ATGGAACCCAT TATCCCCACT GCCAAGTCCA CCCCACACT CATTAGCAA
51 AGCCAAACCCA AGTCTGCCAG TGAGAAGTCA CAGCGCAGCA AGAAGGCCAA
101 GGAGCTGAAG CCAAAGGTGA AGAAGCTCAA GTACCACCAG TACATCCCCC
151 CGGACCAAGAA GCAGGACAGG GGGGCACCCC CCATGGACTC ATCCTACGCC
201 AAGATCCTGC AGCAGCAGCA GCTCTTCTC CAGCTGCAGA TCCTCAACCA
251 GCAGCAGCAG CAGCACCACA ACTACCAGGC CATCCTGCCT GCCCGGCCAA
301 AGTCAGCAGG CGAGGCCCTG GGAAAGCAGC GGACCCCCC AGTACGCAGC
351 CTCTCCACTA CCAATAGCAG CTCCAGCTCG GGCGCCCTG GGCCTGTGG
401 GCTGGCAGCT CAGAACAGCA CCTCACTGAC TGGCAAGCCG GGAGCCCTGC
451 CGGCCAACCT GGACGACATG AAGGTGGCAG AGCTGAAGCA GGAGCTGAAG
501 TTGGCATCAC TGCTGTCTC GGGCACCAA ACTGAGCTGA TTGAGCGCCT
551 TCGAGCCTAT CAAGACCAA TCAGCCCTGT GCCAGGAGCC CCAAGGCCC
601 CTGCGCCGAC CTCTATCCTG CACAAGGCTG GCGAGGTGGT GGTAGCCTTC
651 CCAGCGGCCG GGCTGAGCAC GGGGCCAGCC CTGGTGGCAG CAGGCCCTGC
701 TCCAGCTGAG GTGGTGGTGG CCACGGTGGC CAGCAGTGGG GTGGTGAAGT
751 TTGGCAGCAC GGGCTCCACG CCCCCCGTGT CTCACACCCC CTCGGAGCGC
801 TCACTGTCTA GCACGGGCGA TGAAAACTCC ACCCCCGGGG ACACCTTTGG
851 TGAGATGGTG ACATCACCTC TGACGCAGCT GACCTGCAG GCCTCGCCAC
901 CTCAGATCCT CGTGAAGGAG GAGGGCCCCC GGGCCGGGTC CTGTTGCCCTG
951 AGCCCTGGGG GGGCGGCGGA GCTAGAGGGG CGGCACAAGG ACCAGATGCT
1001 GCAGGAGAAA GACAAGCAGA TCGAGGCGCT GACGCGCATG CTCGGGCAGA
1051 AGCAGCAGCT GGTGGAGCGG CTCAGCTGC AGCTGGAGCA GGAGAAGCGA
1101 GCCCAGCAGC CCGCCCCCGC CCCCCCCCCC CTCGGCACCC CCGTGAAGCA
1151 GGAGAACAGC TTCTCCAGCT GCCAGCTGAG CCAGCAGCCC CTGGGCCCGC
1201 CTCACCCATT CAACCCAGC CTGGCGGCC CAGCCACCAA CCACATAGAC
1251 CTTGTGTGCT TGCCCCCAGG GCCCCCGTCC GTGGTGGTGA AGCAGGAAGC
1301 CTTGCAGCCT GAGCCCCGAG CGGTCCCCGC CCCCCAGTTG CTTCTGGGCG
1351 CTCAGGGCCC CGGCCTATC AAGGGGGTTG CACCTCCAC CCTCATCACC
1401 GACTCCACAG GGACCCACCT TGTCTCACG GTGACCAATA AGAATGCAGA
1451 CAGCCCTGGC CTGTCCAGTG GGAGCCCCCA GCAGCCCTCG TCCCAGCCTG
1501 GCTCTCCAGC GCCTGCCCCC TGTGCCAGA TGGACCTGGA GCACCCACTG
1551 CAGCCCTCTT TTGGGACCCC CACTTCTCTG CTGAAGAAGG AACCACCTGG
1601 CTATGAGGAA GCCATGAGCC AGCAGCCCAA ACAGCAGGAA AATGGTTCTT
1651 CAAGCCAGCA GATGGACGAC CTGTTTGACA TTCTCATTCA GAGCGGAGAA
1701 ATTTACAGCAG ATTTCAAGGA GCCGCCATCC CTGCCAGGGA AGGAGAAGCC
1751 ATCCCCAAG ACAGTCTGTG GGTCCCCCTT GGCAGCACAG CCATCACCTT
1801 CTGCTGAGCT CCCCCAGGCT GCCCCACCTC CTCCAGGCTC ACCCTCCCTC
1851 CCTGGAGCCC TGGAGGACTT CCTGGAGAGC AGCACGGGGC TGCCCTGTCT
1901 GACCAGTGGG CATGACGGGC CAGAGCCCTT TTCCCTCATT GACGACCTCC
1951 ATAGCCAGAT GCTGAGCAGC ACTGCCATCC TGGACCACCC CCGTCAACCC
2001 ATGGACACCT CGGAATTGCA CTTTGTTCCT GAGCCCAGCA GCACCATGGG
2051 CCTGGACCTG GCTGATGGCC ACCTGGACAG CATGGACTGG CTGGAGCTGT
2101 CGTCAGGTGG TCCCGTGTCT AGCCTAGCCC CCCTCAGCAC CACAGCCCCC
2151 AGCCTCTTCT CCACAGACTT CCTCGATGGC CATGATTTCG AGCTGCACTG
2201 GGATTCTCTG TTGTAGCTCT CTGGCTCAAG ACGGGGTGGG GAAGGGGCTG
2251 GGAGCCAGGG TACTCCAATG CGTGGCTCTC CTGCGTGATT CGGCCCTCTC
2301 ACATGGTTGT GAGTCTTGAC AATCACAGCC CTGCTTTT CTCTCCCTG
2351 GGAGGCTAGA ACAGAGAAGC CTTTACTCCT GGTTCAGTGC CACGCAGGGC
2401 AGAGGAGAGC AGCTGTCAAG AAGCAGCCCT GGCTCTCACG CTGGGGTTTT
2451 GGAGACACAG TCAGGGTCAG GGCCATTTCA GCTTGACCTC CTTTTTTGAG
2501 GTCAGGGGGC ACTGTCTGTC TGGCTACAAT TTGGCTAAGG TAGGTGAAGC
2551 CTGGCCAGGC GGGAGGCTTC TCTTGTGACC CAGGGCTGAG ACAGGTTAAG
2601 GGGTGAATCT CTTTCTTTT TCTCCCTGCT TTGCTGTGAA GGGAGAAATT
2651 AGCCTGGGCC TCTACCCCTT ATTCCCTGTG TCTGCCAACC CCAGGATCCC
2701 AAGGCTCCCT GCCATTTTAG TGTCTTGGTG TAGTGTAACC ATTTAGTGGT
2751 TGGTGGCAAC AATTTTATGT ACAGGTGTAT ATACCTCTAT ATTATATATC
2801 GACATACATA TATATTTTGG GGGGGGGGCG GACAGGAGAT GGGTGCAACT
```

```

2851 CCCTCCCATC CTA CTCTCAC AGAAGGGCCT GGATGCAAGG TTACCCCTGA
2901 GCTGTGTGCC ACAGTCTGGT GCCCAGTCTG GCATGCAGCT ACCCAGGCCC
2951 ACCCATCACG TGTGATTGAC ATGTAGGTAC CCTGCCACGG CCTATGCCCC
3001 ACCTGCCCTG CTTCTGGCT CTTATCAGT GCCATGAGGG CAGAGGTGCT
3051 ACCTGGCCTT CCTGCCAGGA GCTCTCCACC CACTCACATT CCGTCCCCGC
3101 CGCCTCACTG CAGCCAGCGT GGCCTTAGGA CAGGAGGAGC TTCGGGCCCCA
3151 GCTTCACCCCT GCGGTGGGGC TGAGGGGTGG CCATCTCCTG CCCTGGGGCC
3201 ACTGGCTTCA CATTCTGGGC TGA CTATAG GGGAGTAGGG GTGGAGTCAC
3251 CAAAACCACT GCTGGGACAA AGATGGGGAA GGTGTGTGAA CTTTTTAAAA
3301 TAAACACAAA AACACAGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAG

```

BLAST Results

Entry HS1042K10 from database EMBL:
 Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2.
 Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2,
 Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP
 domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a
 putative CpG island.
 Score = 7997, P = 0.0e+00, identities = 1617/1645
 7 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 183 bp to 2213 bp; peptide length: 677
 Category: similarity to unknown protein
 Classification: unclassified

```

1 MDSSYAKILQ QQQLFLQLQI LNOQQQOHHN YOAILPAPPK SAGEALGSSG
51 TPPVRSLSST NSSSSSSGAPG PCGLARQNST SLTGKPGALP ANLDDMKVAE
101 LKQELKLRSL PVSGTKTELI ERLRAYQDQI SPVPGAPKAP AATSILHKAG
151 EVVVAFFPAAR LSTGPALVAA GLAPAEVVVA TVASSGVVVF GSTGSTPPVS
201 PTPSERSLLS TGDENSTPGD TFGEMVTSPL TQLTLOASPL QILVKEEGPR
251 AGSCCLSPGG RAELEGRDKD QMLQEKDKQI EALTRMLRQK QQLVERLKLQ
301 LEQEKRAQQP APAPAPLGTP VKQENSFSSC QLSQQPLGPA HPFNPSLAAP
351 ATNHIDPCAV APGPPSVVVK QEALQPEPEP VPAPQLLLGP QGPGLIKGVA
401 PPTLITDSTG THLVLTVIN KADSPGLSSG SPQQPSSQPG SPAPAPSAQM
451 DLEHPLQLPF GTPTSLLKKE PPGYEEAMSQ QPKQENGSS SQQMDLFDI
501 LIQSGEISAD FKEPPSLPGK EKPSPKTVCG SPLAAQSPS AELPQAAPP
551 PGSPSLPGRL EDFLESSTGL PLLTSGHDGP EPLSLIDDLH SQMLSSTAIL
601 DHPPSPMDTS ELHFVPEPSS TMGLDLADGH LDSMDWLELS SGPVLSLAP
651 LSTTAPSLFS TDFLDGHDQ LHWDSCL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22n13, frame 3

TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island., N = 1, Score = 1285, P = 4.9e-131

TREMBL:CEUK06A9_3 gene: "K06A9.1a"; Caenorhabditis elegans cosmid K06A9., N = 2, Score = 149, P = 1.3e-09

TREMBLNEW:SSI132828_1 product: "p210 protein"; Spermatozopsis similis mRNA for p210 protein, partial, N = 1, Score = 171, P = 2.8e-09

>TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC

HSPs :

Query:	435	PSSQPGSPAPAPSAQMDLEHLPLQLFGTPTSLLLKKEPPGYEAMSSQPKQKQENGSSSQOM	494
Sbjct:	1	PSSQPGSPAPAPSAQMDLEHLPLQLFGTPTSLLLKKEPPGYEAMSSQPKQKQENGSSSQOM	60
Query:	495	DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPAELPQAAPPPPGSP	554
Sbjct:	61	DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPAELPQAAPPPPGSP	120
Query:	555	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPSPMDTSELHF	614
Sbjct:	121	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPSPMDTSELHF	180
Query:	615	VPEPSSMTGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	674
Sbjct:	181	VPEPSSMTGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	240
Query:	675	SCL 677	
Sbjct:	241	SCL 243	

Report for DKFZphtes3_22n13.3

```

SEQ      MDSSYAKILQQQQLFLQLQILNQQQQHHNYQAILPAPPKSAGEALGSSGTPVRSLSLT
SEG      . . . . . xxxxxxxxxxxxxxxxxxxxxxx . . . . . xxxxxx
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhcceeceeecccccceeeccccccccceeeccc
COILS    . . . . .
MEM      . . . . .

SEQ      NSSSSSGAGPGCGLARQNSTSLTGKPGALPANLDDMKVAELKQELKLRSLPVS
SEG      xxxxxx
PRD      cccccccccccceeeccccccccccccccccccccchhhhhhhhhhhhhhhhhc
COILS    . . . . .
MEM      . . . . .

SEQ      ERLRAYQDISPVPGAPKAPAATSILHKAGEVVVAFPAARLSTGPALVAAGLAPAE
SEG      . . . . . xxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhcccccccccccceeeceeecccccceccccccccccccceeeeee
COILS    . . . . .
MEM      . . . . . MMMMMMMMMMMMMMMMMMMMMM

SEQ      TVASSGVVKFGSTGSTPPVSPTPSERSLLSTGDENSTPGDTFGEMVTSPLTQLT
SEG      xxxxxxxx . xxxxxxxxxxxxxxxx
PRD      eeeccccccccccccccccccccceeecccccceccccccccceeeccceeeccccc
COILS    . . . . .
MEM      M . . . . .

SEQ      QILVKEEGPRAGSCCLSPGGRAELGRDKDQMLQEKKQIEALTRMLRQKQLVERL
SEG      KQ
PRD      eeeccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    . . . . .
MEM      . . . . . CCCCCCCCCCCCCCCCCCCCCC

SEQ      LEQEKRAQAPAPAPAPLGTVPVKQENSFSSCQLSQQPLGAHPFNPSLAAPATNHD
SEG      PCAV

```

```
SEG      .....xxxxxxxxx.....
PRD      hhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    CCCCCC.....
MEM      .....

SEQ      APGPPSVVVKQEQALQPEPEVPAPQQLLGPGGGLIKGVAPPTLITDSTGTHLVLTVTK
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      NADSPGLSSGSPQPSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLLKKEPPGYEEAMSQ
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      QPKQKENGSSSQMDDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPS
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      AELPQAAPPPPGSPSLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAIL
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      DHPPSPMDTSELHFVPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      TDFLDGHDLQLHWDSC
SEG      .....
PRD      cccccccccccccccccc
COILS    .....
MEM      .....
```

(No Prosite data available for DKFZphtes3_22n13.3)

(No Pfam data available for DKFZphtes3_22n13.3)

DKFZphtes3_23111

group: intracellular transport and trafficking

DKFZphtes3_23111 encodes a novel 186 amino acid protein nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).

Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors (ARFs) to donor membranes, leading to recruitment of cocatomer, bud formation, and eventual vesicle release. ARFs are approximately 20-kDa GTPases that are active with bound GTP and inactive with GDP bound. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and seems to be a novel ARF. It seems to have an important role in vesicular transport and vesicular trafficking.

The new protein can find application in modulating vesicle transport and trafficking in cells.

nearly identical to mouse Arl6, ADP-ribosylation-like factor homolog

start at Bp 15 matches kozak consensus ANNAtgG

Sequenced by LMU

Locus: unknown

Insert length: 717 bp

Poly A stretch at pos. 689, no polyadenylation signal found

```
1 ATTTGAATCA CATTATGGGA TTGCTAGACA GACTTTCAGT CTTGCTTGGC
51 CTGAAGAAGA AGGAGGTTC TGTTTTGTGC CTTGGGCTAG ATAATAGTGG
101 CAAAACGACG ATCATTAAAC AACTTAAACC TTCAAATGCT CAATCTCAAA
151 ATATCCTTCC AACAAATAGGA TTCAGCATAG AGAAATTCAA ATCATCCAGT
201 TTGTCAATTA CAGTGTTTGA CATGTCAGGT CAAGGAAGAT ACAGAAATCT
251 CTGGGAACAC TATTATAAAG AAGGCCAAGC TATTATTTT GTCATTTGATA
301 GTAGTGATAG ATTAAGAATG GTTGTGGCCA AAGAAGAAGT CGATACTCTT
351 CTGAATCATC CAGATATTAA ACACCGTCGA ATTCCAATCT TATTCTTTGC
401 AAATAAAATG GATCTTAGAG ATGCAGTGAC ATCTGTAAAA GTGTCTCAGT
451 TGCTGTGTTT AGAGAATATC AAAGATAAAC CCTGGCATAT TTGTGCTAGT
501 GATGCCATAA AAGGAGAAGG CTTGCAAGAA GGTGTAGACT GGCTTCAAGA
551 TCAGATCCAG ACTGTGAAGA CATGAAAAGA TAATAGTTGG AAACCTCAGC
601 AATTTTCAAT TCAAGGAATC TATCTAAGAC AAATAGAATA CATTTTGTAA
651 AAGATGTTTA TGCATCAAAA AATATAATT TCTGCTTGCA AAAAAAAAAA
701 AAAAAAAAAA AAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 15 bp to 572 bp; peptide length: 186
Category: strong similarity to known protein
Classification: Intracellular transport and traffic
Prosite motifs: ATP_GTP_A (24-32)

```
1 MGLLDRLSVL LGLKKKEVHV LCLGLDNSGK TTIINKLKPS NAQSQNILPT
51 IGFSIEKFKS SLSFTVFDM SGQGRYRNW EHYKQEQAI IFVIDSSDRL
101 RMVVAKEELD TLLNHPDIKH RRIPILFFAN KMDLRDAVTS VKVSQLLCLE
151 NIKDKPWHIC ASDAIKGEGL QEGVDWLQDQ IQTVKT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23111, frame 3

TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds., N = 1, Score = 923, P = 1.1e-92

TREMBL:CEC38D4.5 gene: "C38D4.8"; Caenorhabditis elegans cosmid C38D4, N = 1, Score = 418, P = 3.6e-39

PIR:S66337 ADP-ribosylation factor 1 - Chlamydomonas reinhardtii, N = 1, Score = 373, P = 2.1e-34

SWISSPROT:ARF1_CHLRE ADP-RIBOSYLATION FACTOR 1., N = 1, Score = 372, P = 2.7e-34

>TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds.
Length = 186

HSPs:

Score = 923 (138.5 bits), Expect = 1.1e-92, P = 1.1e-92
Identities = 178/186 (95%), Positives = 184/186 (98%)

Query: 1 MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS 60
MGLLDRLS LLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQ+I+PTIGFSIEKFKS
Sbjct: 1 MGLLDRLSGLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQDIVPTIGFSIEKFKS 60

Query: 61 SSLSFTVFDMMSGQGRYRNLEWHYYKQGAIIFVIDSSDLRMVVAKEELDTLLNHPDIKH 120
SSLSFTVFDMMSGQGRYRNLEWHYYK+GQAIIFVIDSSD+LRMVVAKEELDTLLNHPDIKH
Sbjct: 61 SSLSFTVFDMMSGQGRYRNLEWHYYKDGQAIIFVIDSSDKLRMVVAKEELDTLLNHPDIKH 120

Query: 121 RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180
RRIPILFFANKMDLRD+VTSVKVSQLLCLE+IKDKPWHICASDAIKGEGLQEGVDWLQDQ
Sbjct: 121 RRIPILFFANKMDLRDSVTSVKVSQLLCLESIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180

Query: 181 IQTVKT 186
IQ VKT
Sbjct: 181 IQAVKT 186

Pedant information for DKFZphtes3_23111, frame 3

Report for DKFZphtes3_23111.3

[LENGTH] 186
[MW] 21097.69
[PI] 8.72
[HOMOL] TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds. 4e-94

[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YDL192w] 1e-36
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 1e-36
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL192w] 1e-36

[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL137w] 2e-36

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YBR164c] 2e-32
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 2e-32
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 4e-19
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 4e-19
[FUNCAT] x general function prediction [M. jannaschii, MJ1339] 2e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 4e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHR005c] 4e-05

[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YHR005c] 4e-05
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YKR014c] 2e-04
[FUNCAT] 08.19 cellular import [S. cerevisiae, YKR014c] 2e-04
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YKR014c] 2e-04

[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-04

[BLOCKS] BL01288C
[BLOCKS] BL01020C SAR1 family proteins
[BLOCKS] BL01019C ADP-ribosylation factors family proteins


```

[BLOCKS]      BL01019B ADP-ribosylation factors family proteins
[BLOCKS]      BL01019A ADP-ribosylation factors family proteins
[SCOP]         dias3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domain 2e-45
[SCOP]         dlmhl_ 3.29.1.4.2 Rac1 [Human (Homo sapiens)] 2e-46
[SCOP]         d5p21_ 3.29.1.4.1 cH-p21 Ras protein [human (Homo sapiens)] 5e-37
[SCOP]         dihura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) [human (Homo sapiens)] 4e-61
[SCOP]         dla2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) [Do] 4e-33
[PIRKW]        glycoprotein 2e-33
[PIRKW]        monomer 3e-31
[PIRKW]        P-loop 2e-35
[PIRKW]        lipoprotein 2e-33
[PIRKW]        GTP binding 2e-35
[SUPFAM]       ADP-ribosylation factor 2e-35
[PROSITE]      ATP_GTP_A 1
[PFAM]         ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW]           Alpha_Beta
[KW]           3D
[KW]           LOW_COMPLEXITY 5.91 %

```

```

SEQ      MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS
SEG      ..xxxxxxxxxxxxx
lhurA    .....CCCCEEEEETTTTCHHHHHHHHCCCCEEEE--EEETEEEEEEEE

```

```

SEQ      SLSFTVFDMSGQGRYRNLEWHYYKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH
SEG      .....
lhurA    TTEEEEEETTTTTTCCCHHHHHHCEEEEEETTTTHHHHHHHHHHHHHHTTT--

```

```

SEQ      RRIPILFFANKMDLRDAVTSVKVSQLLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ
SEG      .....
lhurA    TTTEEEEEETTTTTTCCCHHHHHHCGGGTTTTCEEEECBTBTBHHHHHHHHHHHH

```

```

SEQ      IQTVKT
SEG      .....
lhurA    HHHHC.

```

Prosite for DKFZphtes3_23111.3

```

PS00017      24->32  ATP_GTP_A      PDOC00017

```

Pfam for DKFZphtes3_23111.3

```

HMM_NAME      ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
HMM            *GMgWfsIFrkmWGLWNKEMRILMLGLDNAGKTTIILYMLKlgE..IVTTI
               MG++ ++ ++GL +KE+++L LGLDN+GKTTI+++LK+ ++
Query          1  -MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNIL 48
HMM            PTIGFNVETVeYKNIKFNVVDVGGQdsIRPYWRHYYPNTDGLIIVVDSaD
               PTIGF +E+ + ++F+V+D GQ + R +W HYY + ++II+V+DS+D
Query          49 PTIGFSIEKFSSLSFTVFDMSGQGRYRNLEWHYYKEGQAIIFVIDSSD 98
HMM            RDRMeEaKqELHaMLNEEEL..rDAPILIFANKQDLPgAMSesEIREaLG
               R RM AK+EL+ +LN+ ++ R+ P+L FANK DL++A+++ +++ +L
Query          99 RLRMVVAKEELDTLLNHPDIKHRRIPILFFANKMDLRDAVTSVKVSQLLC 148
HMM            LHeIRCnRPWYIQMCCAvtGEGLYEGMDWLSNYInkrkK*
               L++I+ + PW+I +++A++GEG+EG DWL ++I+ K
Query          149 LENIK-DKPWHICASDAIKGEGLQEGVDWLQDQIQTVKT 186

```

DKFZphtes3_23n19

group: testes derived

DKFZphtes3_23n19 encodes a novel 387 amino acid protein with similarity to rat protein kinase C-interacting RBCC protein 1.

The novel protein contains not the RING-B box-coiled coil (RBCC) motif of RBCC protein 1, and thus is not a member of this subgroup of RING finger proteins.
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp

Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
1 CGGAGACCCT CGGGCCGTGT CCATTGTGG GCAAAGCCAG CGGGGCAGGC
51 TTGGCCAGAG TGCACCACTC GCGCCCGTCC CAGGCCCGAC GCTCTGGGCG
101 CGCCCCGAAC CCCAGGTTTC GGGCCCGTGT TTCCGACCGG CGGAGGGGGC
151 TCAGCGGGCC GATCCACGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
201 GACCGGAGAT GCGGCCGCCA GCGGGCGGG GCGCGCGGGC GGCCTCGGAC
251 TTGGGCTCCG CCGCAGTGCT CTGGCTGTG CACGCCCGGG TGAGGCGGCT
301 GGGCGCCGGG CCAGACGCGG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
351 CGGACCTTGA GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
401 CTTGGGGCGG TTAATTGGA GTGGCCCTG GAGTCAGTTT CCTACACCAT
451 CCGAGGCCCC ACCCAGCAGC AGCTACAGCC TCCACCAGGA GGGCCTGGAA
501 CCCTCAGCCT GCACTTCCTC AACCTCAGG AAGCTCAGCG GTGGGCAGTC
551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
601 ACCACCCAGC TTGGGCCCCG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
651 CGGAAGCCTC CACACTCAAG GGCCTCCAC CTGAGGCAGA TCTTCCTAGG
701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
801 TGGCCACGCA TCGTGTGGCC CTGAGTGTTC AGCTTCAGGA GGCCTGCTTC
851 CCACCTGGCC CCATCAGGCT GCAGGTCACA CTTGAAGACG CTGCCTCTGC
901 CGCATCCGCC GCGTCTCTG CACACGTTGC CCTGCAGGTC CACCCCACTC
951 GCACTGTTGC AGCTCTCCAG GAGCAGGTGT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCTGTGTG TGCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCGGCAGGA TGGGGACCTT GCTTTCCTCT
1101 ACTTGCTGTC AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCCAGA AGATGGACGG GGAACCTGGA CGCTTGTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCAGCCTG CCCAGTCCAC
1251 TCCAGCCAG CTGGTCTGTG CCTTCTGCA CCTTCATCAA TGCCCCAGAC
1301 CGCCCTGGCT GTGAGATGTG TAGCACCAG AGGCCCTGCA CTGGGACCC
1351 CCTTGCTGCA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCTTCCC TCACAAGTCC GACATCTCCA GGGCCCACT GAACTCCGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGTTG GGGGAAGGG
1501 CCACAAAATG AAACCATTA AGACCTTAA GAGCCAAAAA AAAAAAAA
1551 AAAAAAAA AAAAAAAA AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387

Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAAEQ LRRQLSADP
51 ERPGFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTLS
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLLPSPPEA
151 STLKGPPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLQA
201 HRVALSVQLQ EACFPFGPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLFPFS LGLPPGPQPA ASSLPSPQLP
351 SWSCPSTFI NAPDRPGCEM CSTQRPCTWD PLAAAST

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1,
 Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
 complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
 HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
 = 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
 musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
 = 367, P = 9.3e-34

>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
 UbcM4 interacting protein 28 mRNA, complete cds.
 Length = 498

HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

```

Query: 175 LAGSLARAIAGGDEKGAQAQVAVLAQHRVALSVQLQEACFPFGPIRLQVTLEDAASAASA 234
      +A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA
Sbjct: 1 MALSLARAVAGGDEQAIAKYATWLAQVRVPLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDGP 294
      + + L V P TVA+L++ VF + GFPP++Q+WW+G+ L + +L S+G+R++GD
Sbjct: 57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQWVVGQRLARDQETLHSHGIRRNQDG 115

Query: 295 AFLYLILLSAPREAPATGSPQHPQK----MDGELG--RLFPPSLG-LPPG-PQPAASSLP 345
      A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P
Sbjct: 116 AYLYLLSARN----TSLNPQELQRQRRLMEDLGFKDLTLQSRGPLEPVLKPKRTNQEP 171

Query: 346 -----SPLQP--SWSCPSTFINAPDRPGCEMCSTQRPCTW 379
      +P P W CP CTFIN P RPGCEMC RP T+
Sbjct: 172 GQPDAAAPESPPVGWQCQPGCTFINKPTRPGCEMCCRARPETY 212

```

Pedant information for DKFZphtes3_23n19, frame 2

Report for DKFZphtes3_23n19.2

```

[LENGTH]      387
[MW]           39949.29
[pI]           5.53
[HOMOL]        TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]       BL00578B
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY 17.57 %

```

```

SEQ  MAPPAGGAAAAASDLGSAAVLLAVHAAVRPLGAGPDAAEQQLRRQLSADPERPGFRLEL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhccccccccceee

```

```
SEQ      LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPGGPGTLSLHFLNPQEAAQRWAVLVRGAT
SEG
PRD      cccccccceeecccceeeeeecccccccccccccccccceeeecccchhhhhhheeeccc

SEQ      VEGONGSKSNSPPALGPEACPVSPLSPPEASTLKGPPEADLPSPGNLTEREELAGSLA
SEG
PRD      .....xxxxxxxxxxxxx.....
eccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ      RAIAGGDEKGAAQVAVALQAHRVALSVQLQEACFPPIRLQVTLEDAASAASASSAHV
SEG
PRD      hhhhhccccchhhhhhhhhhhhhhhhhhhhhccccccccccccceecchhhhhhhhhhhhee

SEQ      ALQVHPHTVAALQEQQVFSELGFPPAVQRWVIGRCLCVPERSLASVYGRQDGDPAFLYLL
SEG
PRD      .....
eeccccchhhhhhhhhhhhhccccccchhhhhhhhhhhccccccccccccceeeec

SEQ      SAPREAPATGPSQPHQKMDGELGRLFPPSLGLPPGPQAASSLPSPLOPSWSPCSTFI
SEG
PRD      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
ccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccee

SEQ      NAPDRPGCEMCSTQRCTWDPLAAAST
SEG
PRD      .....
ccccccccccccccccccccccccceeeccc
```

(No Prosite data available for DKFZphtes3_23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp
Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

1	CGGAGACCTT	CGGGCCGTGT	CCATTTGTGG	GCAAAGCCAG	CGGGGACGCG
51	TGTTGCCACAG	TGCACCACTT	GGCGCCCGTTC	CAGGCCCCAG	GCTCTTGGGGC
101	CGCCCGGAAC	CGGAGCTTGC	GGCCCGCTGC	TTTCCACAGT	CGGAGGGGGCG
151	TACGGCGGAC	GATCCCAACG	AAGCGCGCTC	GAGGGGGTGG	GACCCCGGGCG
201	CACCGGAGAT	GGCGCGCCGA	CGGGCGGGGG	CGGCGCGGGC	CGCTCTCGAAC
251	TTGGGCTCCG	CCGCAGTGTCT	CTTGCTGTGT	CACGCCCGGG	TGAGGCGCGG
301	GGGCGCCGGG	CAGACGCGCC	AGGCACAGCT	CGGGAGGCTG	CAGCTGAGCG
351	CGGACCTTGA	TAGGCGCTGG	CGTTTCCGGC	TGGAGCTGCT	GGGCGCGGGA
401	CTGTGGGCGG	TTAATTGGA	GTGGCCCGCT	GAGTCAGATT	CCTACACCAT
451	CCGAGGCGCC	ACCCAGCAGC	AGCTACAGCG	TATCCAGAGGA	GGGCTCTGAA
501	CCTCAGCCTT	GCACTTCCCT	AACCTCAGG	AAGTCCAGCG	GTGGGCGAGTC
551	CTAGTCCGAG	GTGCCACCTG	GGAAGGACAT	AATGGCAGCA	AGAGCAACTCT
601	ACCCACGAGC	TTGGGCCCAG	AAGCATGCC	TGTCTTCCCT	CCCAGTCCCC
651	CGGAAGCCTC	CACACTCAAG	GGCCCTCCAC	CTGAGGAGCA	TCTCTCTAGG
701	AGCCTCTGAA	ACTTGACAGA	GAGAGAAGAG	CTGGCAGGGA	GCGCTGGCCCG
751	GCGTATTGCA	GGTGGAGACG	AGAAGGGGGG	AGCCCAAGTC	GCAGCCGCTCG
801	TGGCCACAGA	TCGTGTGGCC	CTAGTGTGTC	AGCTTCAGGA	GGCCTGCTTC
851	CCACCTTGCC	CCATCAGGCT	GCAGGTACAC	CTTGAAGACG	CTGCTCTTGC
901	CGCATTCGCC	GGTCTCTCTG	CACAGCTTGC	CTCAGAGTGC	CACCCCCCAT
951	GCACTGTGTG	AGCTCTCCAG	GAGCAGGTGT	TCTAGAGCTC	CGGTTTCCCG
1001	CACGCGCTGC	AAGCTGTGGT	CATCGGACGC	TGCGTGTGTC	TGCTTGAAGC
1051	GAGCCTTGCT	CTTACGGGGT	TTCCGAGGA	TGGGGAGGTT	GCTTTCTCTCT
1101	ACTTGTCTGT	AGCTCCTCGA	GAAGCCCCAG	CACAGGACC	TAGCCCTCAG
1151	CACCCCCAGA	AGATGAGACG	GGAACTTGGA	CGCTTGTTC	CCCCATCAT
1201	GGGGCTATCC	CAGAGCCCCC	AGCCAGTGTG	CTCCAGCGTG	CCAGTCCAC
1251	TCCAGCGGAG	CTGGTCTGTG	CCTTCTGSCA	CCTCTACAA	TGCCCCAGAC
1301	CGCCTCCGCT	GTGAGATGTG	TAGCACCAGC	AGGCCCTGCA	TTTGGGACCG
1351	CCTTGCTGCA	GTTTCCACCT	AGCAGCCACC	AGAGGTTACA	AGGGGAGAGT
1401	GGCCTTCTCC	TCACAAGTCC	GACATCTCCA	GGCCCCCAT	GAACTCCGGG
1451	GACCTCTACT	GACTGCTTGC	TGGGACAGTC	ACACGGGTTG	GGGGGAAGGG
1501	CACCAAAAT	AAACCATTA	AGACCTTTAA	GAGGCCAAAA	AAAAAAAATA
1551	AAAAAAAATA	AAAAAAAATA	AAAAAAAAG		

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387
Category: similarity to known protein
Classification: Cell signaling/communication

```
1 MAPPAGGAAA AASDLGSAV LLAVHAAVRP LGAGPDAAEQ LRLQLSADP
51 ERPGRFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTSL
101 LHFLNPQEAQ RWAFLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLAQ
201 HRVALSVQLQ EACFPPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLEFPPS LGLPPGPQPA ASSLPSPLPQ
351 SWSCPSCTFI NAPDRPGCEM CSTQRPCTWD PLAAAST
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCK protein 1 - rat, N = 1,
Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
= 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
= 367, P = 9.3e-34

>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds.
Length = 498

HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
Identities = 95/212 (44%), Positives = 129/212 (60%)

Query: 175 LAGSLARAIAGGDEKGAAQVAAVLAQHHRVALSVQLQEACFPPIRLQVTLEDAASAASA 234
+A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA
Sbjct: 1 MALSLARAVAGGDEQAIAKYATWLAEQVRPLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGD 294
+ + L V P TVA+L++ VF + GFPP++Q+VV+G+ L + +L S+G+R++GD
Sbjct: 57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQVVGQRLARDQETLHSHGIRRNQD 115

Query: 295 AFLYLLSAPREAPATGPSPOHPQK-----MDGELG--RLFPPLSG-LPPG-PQPAASSLP 345
A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P
Sbjct: 116 AYLYLLSARN----TSLNPQELQRQRLRMLEDLGFKDLTLQSRGPLEPVLKPKRTNQEP 171

Query: 346 -----SPLQF--SWSCPSCTFINAPDRPGCEMCSTQRPCTW 379
+P P W CP CTFIN P RPGCEMC RP T+
Sbjct: 172 GQPDAAPESPVVGWQCPCGCTFINKPTRPGCEMCCRARPETY 212

Pedant information for DKFZphtes3_23n19, frame 2

```
[LENGTH]          387
[MW]              39949.29
[pI]              5.53
[HOMOL]           TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]          BL00578B
[KW]              Alpha_Beta
[KW]              LOW COMPLEXITY      17.57 %
```

```
SEQ      MAPPAGGAAAAAASDLGSAVLLAVHA AVRPLGAGPD AEALRRQLQSADPERPGRFRLEL
SEG      .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```

SEQ      LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPPGPGTSLHFLNPQEAQRWAVLVRGAT
SEG      .....
PRD      cccccccceccccceeeeeccccccccccccccccccccceeeeecccchhhhhheeeccce

```

```
SEQ      VEGQNGSKSNSPPALGPEACPVS LSPPEASTLKGPPEADLPRSPGNLTEREELAGSLA
SEG      .....
PRD      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeechhhhhhhhhhhh
```

SEQ RAIAGGDEKGAAQVAVLAQHRVALSVQLQEACFPGPRLQVTLEDAASAASAASHAV
SEGxxxxxxxxxxxxx..
PRD hhhccccchhhhhhhhhhhhhhhhhcccccceccchhhhhhhhhhhee

SEQ ALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDPAFLYLL
SEG
PRD eeccccchhhhhhhhhhhhhccccchhhhhhhhhhhcccccccccccccccccccc

```
SEQ  SAPREAPATGPSQHPQKMDGELGRLFPSPSLGLPPGPQPAASSLPSLQPSWSWCPSTFI
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...
PRD  cccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccc
```

```

SEQ      NAPDRPGCEMCSTQRPCTWDPLAAAST
SEG      .....
PRD      cccccccccccccccccccccceeeccc

```

(No Prosite data available for DKFZphtes3_23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

DKFZphtes3_26g22

group: intracellular transport/trafficking

DKFZphtes3_26g22 encodes a novel 898 amino acid protein with similarity to kinesins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport. It is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain contains a large globular N-terminal domain which is responsible for the motor activity of kinesin, which is known to hydrolyze ATP and to bind and move on microtubules. Several proteins involved in chromosome segregation and cell division contain this motor domain, such as drosophila claret segregational protein (ncd), Drosophila kinesin-like protein (nod), human CENP-E and human mitotic kinesin-like protein-1 (MKLP-1). The novel protein is a new kinesin like prottein.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

strong similarity to kinesins

Sequenced by EMBL

Locus: unknown

Insert length: 3032 bp

No poly A stretch found, no polyadenylation signal found

```
1 CTGAAGCGCT GGGAGGCGGA CATTAAAGTG AAGTGGTTGC GGTAACCTGG
51 CCTGGGCGCTG AAGTGAGTGA GAGGCACATG AAGAGAAGTA TTCAAGTATT
101 TATACAGATA GGAATCAAGA TAATCAACAA TGCTGTGCAC TGAGGAAGAC
151 CTGTGCCACC ATATGAAAAGT AGTAGTTTCGT GTACGTCCTGG AAAACACTAA
201 ACAGAAAAGCA GCTGGATTTC ATAAAGTGGT TCATGTTGTG GATAAACATA
251 TCCTAGTTTT TGATCCCAA CAAGAAGAAG TCAGTTTTTT CCATGGAAAG
301 AAAACTACAA ATCAAAATGT TATAAAGAAA CAAAATAAGG ATCTTAAATT
351 TGTATTTGAT GCTGTTTTTG ATGAAACGTC AACTCAGTCA GAAGTTTTTG
401 AACACACTAC TAAGCCAAAT CTTCGTAGTT TTTTGAATGG ATATAATTGC
451 ACAGTACTTG CCTATGGTGC CACTGGTSGT GGGAAAGACCC ACACATGCT
501 AGGATCAGCT GATGAACCTG GAGTGATGTA TCTAACAAATG TTACACCTTT
551 ACAAATGCAT GGATGAGATT AAAGAAGAGA AAATATGTAG TACTGCAGTT
601 TCATATCTGG AGGTATATAA TGAACAGATT CGTGATCTCT TAGTAAATTC
651 AGGGCCACTT GCTGTCCGGG AAGATACCCA AAAAGGGGTG GTCGTTTCATG
701 GACTTACTTT ACACCAGCCC AAATCCTCAG AAGAAATTTT ACATTTATTG
751 GATAATGGAA AAAAAACAG GACACAACAT CCCACTGATA TGAATGCCAC
801 ATCTTCTCGT TCTCATGCTG TTTTCCAAAT TTACTTGCGA CAACAAGACA
851 AAACAGCAAG TATCAATCAA AATGTCCGTA TTGCCAAGAT GTCACCTATT
901 GACCTGGCAG GATCTGAGCG AGCAAGTACT TCCGGTGCTA AGGGGACCCG
951 ATTTGTAGAA GGCACAAATA TTAATAGATC ACTTTTAGCT CTTGGGAATG
1001 TCATCAATGC CTTAGCAGAT TCAAAGAGAA AGAATCAGCA TATCCCTTAC
1051 AGAAATAGTA AGCTTACTCG CTTGTTAAAG GATTCTCTTG GAGGAAACTG
1101 TCAAACATATA ATGATAGCTG CTGTTAGTCC TTCTCTGTA TTCTACGATG
1151 ACACATATAA CACTCTTAAG TATGCTAACC GGGCAAAAGG CATTAAATCT
1201 TCTTTGAAGA GCAATGTTCT TAATGTCAAT AATCATATAA CTCAATATGT
1251 AAAGATCTGT AATGAGCAGA AGGCAGAGAT TTTATTGTGA AAAGAAAAAC
1301 TAAAGCCTTA TGAAGAACAG AAAGCCTTCA CTAATGAAAA TGACCAAGCA
1351 AAGTTAATGA TTTCAAACCC TCAGGAAAAA GAAATCGAAA GGTTTCAAGA
1401 AATCCTGAAC TGCTTGTTC AGAATCGAGA AGAAATTAGA CAAGAATATC
1451 TGAAGTTGGA AATGTTACTT AAAGAAAATG AACTTAAATC ATTCTACCAA
1501 CAACAGTGCC ATAAACAAAT AGAAATGATG TGTCTGAAG ACAAAAGTAGA
1551 AAAGGCCACT GGAACACGAG ATCATAGACT TGCAATGTTG AAAACTCGTC
1601 GCTCCTACCT GGAGAAAAGG AGGGAGGAGG AATTGAAGCA ATTTGATGAG
1651 AATACTAATT GGCTCCATCG TGTCGAAAAA GAAATGGGAC TCTTAAGTCA
1701 AAACGGTCAT ATTCCAAAGG AACTCAAGAA AGATCTTCAT TGTCACCATT
1751 TGCACCTCCA GAACAAAGAT TTGAAAGCAC AAATTAGACA TATGATGGAT
1801 CTAGCTTGTC TTCAGGAACA GCAACACAGG CAGACTGAAG CAGTATTGAA
1851 TGCTTTACTT CCAACCCATA GAAAACAATA TTGCACATTA AAAGAAGCCG
1901 GCCTGTCAAA TGCTGCTTTT GAATCTGACT TCAAAGAGAT CGAACATTTG
1951 GTAGAGAGGA AAAAAGTGGT AGTTGGGCT GACCAAACTG CCGAACACC
2001 AAAGCAAAAC GATCTACCAG GGATTTCTGT TCTTATGACC TTTCCACAAC
2051 TTGGACAGT TCAGCCTATT CTTGTTGCT CATCTTCAGG TGGAACTAAT
2101 CTGGTTAAGA TTCCTACAGA AAAAAGAACT CGGAGAAAAC TAATGCCATC
2151 TCCCTTGAAA GGACAGCATA CTCTAAAGTC TCCACCATCT CAAAGTGTGC
2201 AGCTCAATGA TTCTCTTAGC AAAGAACTTC AGCCTATTGT ATATACACCA
2251 GAAGACTGTA GAAAAGCTTT TCAAAATCCG TCTACAGTAA CCTTAATGAA
2301 ACCATCTACA TTTACTACAA GTTTTCAGGC TATCAGCTCA AACATAAACA
2351 GTGATAATTG TCTGAAAAATG TTGTTGTAAG TAGCTATCCC TCATAATAGA
```

```

2401 AGAAAAGAAT GTGGACAGGA GGAATTGGAC TCTACATTTA CTATATGTGA
2451 AGACATCAAG AGCTCGAAGT GTAAATTACC CGAACAAGAA TCACTACCAA
2501 ATGATAACAA AGACATTTTA CAACGGGCTTG ATCCTTCTTC ATTCTCAACT
2551 AAGCATTCTA TGCCTGTACC AAGCATGGTG CCATCCTACA TGGCAATGAC
2601 TACTGCTGCC AAAAGGAAAC GGAATTAAC AAGTTCTACA TCAAAACAGTT
2651 CGTTAACTGC AGACGTAAAT TCTGGATTG CCAAACGTGT TCGACAAGAT
2701 AATTCAAGTG AGAAGCACTT ACAAGAAAAC AAACCAACAA TGGAACATAA
2751 AAGAAACATC TGTAAATATA ATCCAAGCAT GGTAGAAAAA TTTGGAAGAA
2801 ATATTTCAAA AGGAAATCTA AGATAAATCA CTTCAAAACC AAGCAAAATG
2851 AAGTTGATCA AATCTGCTTT TCAAAGTTA TCAATACCTT TCAAAAAATA
2901 TATTTAAAT CTTTGAAAGA AGACCCATCT TAAAGCTAAG TTTACCCAAG
2951 TACTTTCAGC AAGCAGAAAA ATGAAACTCT TTGTTTCTT CTTTGTGTT
3001 CTAATAAAAT AAAATTTCAA AAGAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 130 bp to 2823 bp; peptide length: 898
 Category: strong similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (113-121)
 KINESIN_MOTOR_DOMAIN1 (252-264)

```

1 MSVTEEDLCH HMKVVVRVRF ENTKEKLAGF HKVVHVVDKH ILVFDPKQEE
51 VSFFHGKKT NQNVIKKQNK DLKFVFDVAF DETSTQSEVF EHTTKPILRS
101 FLNGYNCTVL AYGATGAGKT HTMLGSADEP GVMYLTMLHL YKCMDEIKKEE
151 KICSTAVSYL EVYNEQIRDL LVNSGFLAVR EDTQKGVVVH GLTLHQPKSS
201 EEILHLLDNG NKNRTQHPTD MNATSSRSHA VFQIYLRQOD KTASINQNVF
251 IAKMSLIDLA GSERASTSGA KGRTRFVEGTN INRSLLALGN VINALADSKR
301 KNQHIFPYRN KLTRLKDSL GGNCQTIMIA AVSPSSVFYD DTYNLTLYAN
351 RAKDIKSSLK SNVLNVNNHI TQYVKICNEQ KAEILLLEK LKAYEEQKAF
401 TNENDQAKLM ISNPQEKIE RFQEIILNCLF QNREEIRQEQ LKLEMLLKEN
451 ELKSFYQQQC HKQIEMMCSE DKVEKATGKR DHRRLAMKTR RSYLEKRREE
501 ELKQFEDENTN WLHRVEKEMG LLSQNGHIPK ELKKDLHCHH LHLQNKDLKA
551 QIRHMDLAC LQEQHQRTQ AVLNALLPTL RKQYCTLKEA GLSNAAFESD
601 FKEIEHLVER KVVVWADQT AEQPKQNDLP GISVLMTEPQ LGPVQPIPC
651 SSSGGTNLVK IPTEKRRRK LMPSPKQGH TLKSPSPQSV QLNDSLKSEL
701 QPIVYTPEDC RKAFQNPSTV TLMKPSSTFT SFQAISSNIN SDNCLKMLCE
751 VAIHPNRRKE CGQEDLDSTF TICEKIKSSK CKLPEQESLP NDNKDILQRL
801 DPSSFSTKHS MPVPSMVPSTV MAMTTAAKRK RKLTSSTNS SLTADVNSGF
851 AKRVQDNSS EKHLQENKPT MEHKNRICKI NPSMVRKEGR NISKGNLR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_26g22, frame 1

SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13., N = 3,
 Score = 874, P = 9e-93

TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila
 melanogaster kinesin like protein 67a mRNA, complete cds., N = 1, Score
 = 880, P = 4.2e-88

TREMBL:SPBC649_1 gene: "SPBC649.01c"; product: "putative kinesin-like
 protein"; S.pombe chromosome II cosmid c649., N = 3, Score = 814, P =
 9.8e-86

PIR:S64238 kinesin-related protein KIP3 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 802, P = 2.5e-83

>TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila

melanogaster kinesin like protein 67a mRNA, complete cds.
Length = 814

HSPs:

Score = 880 (132.0 bits), Expect = 4.2e-88, P = 4.2e-88
Identities = 181/345 (52%), Positives = 238/345 (68%)

```
Query: 11 HMKVVVRVRPENTKEAAGFHKVVHVVDKHLVFDPKQEEVSFF-HGKKTNTQNVIKKQN 69
      ++KV VRVRP N +E      ++ V+D+ L+FDP +E+ FF G K +++ K+ N
Sbjct: 8 NIKVAVRVRPYNVRELEQKQRSIIKVMDSALLFDPDEEDDEFFFGAKQPYRDITKRMN 67

Query: 70 KDLKFVFDVAFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGATGAGKTHMLGSADE 129
      K L FD VFD ++ ++FE T P++ + LNGYNC+V YGATGAGKT TMLGS
Sbjct: 68 KKLTMEFDRVFDIDNSNQDLFEECTAPLVDAVLNGYNCVVFYVYATGAGKTFTMLGSEAH 127

Query: 130 PGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQKGVVV 189
      PG+ YLTM L+ + + + VSYLEVYNE + +LL SGPL +RED GVVV
Sbjct: 128 PGLTYLTMQDLFDKIQAQSDVRKFDVGSYLEVYNEHVMNLLTKSGFLKREDNN-GVVV 186

Query: 190 HGLTLHQPKSSEIHLHLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQDKTASINQNV 249
      GL L S+EE+L +L GN +RTQHPTD NA SSRSHA+Q++++R ++ + V
Sbjct: 187 SGLCLTPIYSAEELLRLMLGNSHRTQHPTDANAESSRSHAIFQVHIRITERKTDTKRTV 246

Query: 250 RIAKMSLIDLAGSERASTSGAGKTRFEVGTNINRSLALGNVINALADSKRKNQHIPYRN 309
      K+S+IDLASERA+++ G RF EG +IN+SLLALGN IN LAD + HIPYR+
Sbjct: 247 ---KLSMIDLAGSERAASTKGIGVRFKEGASINKSLLALGNCINKLADGLK---HIPYRD 300

Query: 310 SKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDI 355
      S LTR+LKDSLGGNC+T+M+A VS SS+ Y+DTYNTLKYA+RAK I
Sbjct: 301 SNLTRILKDSLGGNCRTLMVANVSMSSLTEDTYNTLKYASRAKKI 346
```

Pedant information for DKFZphtes3_26g22, frame 1

Report for DKFZphtes3_26g22.1

```
[LENGTH] 898
[MW] 102281.63
[pI] 9.09
[DOMOL] SWISSPROT:YB3D SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13. 3e-97
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKL079w]
4e-28
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 1e-117
[SCOP] d3kar_ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyces 1e-112
[PIRKW] nucleus 6e-87
[PIRKW] heterodimer 4e-68
[PIRKW] DNA binding 9e-60
[PIRKW] heterotetramer 2e-54
[PIRKW] mitosis 9e-60
[PIRKW] microtubule binding 4e-68
[PIRKW] ATP 6e-87
[PIRKW] phosphoprotein 5e-59
[PIRKW] heterotrimer 4e-68
[PIRKW] purine nucleotide binding 1e-26
[PIRKW] P-loop 6e-87
[PIRKW] coiled coil 4e-68
[PIRKW] heptad repeat 3e-62
[PIRKW] methylated amino acid 2e-54
[PIRKW] hydrolase 2e-54
[PIRKW] GTP binding 1e-60
```

```

[PIRKY]      cell division 5e-57
[SUPFAM]     kinesin-related protein KIP1 3e-50
[SUPFAM]     kinesin-related protein CIN8 7e-33
[SUPFAM]     kinesin heavy chain 2e-54
[SUPFAM]     suppressor protein SMY1 1e-26
[SUPFAM]     kinesin-related protein KIF3 4e-68
[SUPFAM]     kinesin-related protein KIF2 1e-46
[SUPFAM]     kinesin-related protein unc-104 7e-60
[SUPFAM]     unassigned kinesin-related proteins 6e-87
[SUPFAM]     centromere protein E 3e-54
[SUPFAM]     kinesin-related protein KLP61F 5e-57
[SUPFAM]     kinesin-related protein MKLP-1 2e-28
[SUPFAM]     pleckstrin repeat homology 7e-60
[SUPFAM]     kinesin-related protein KIF1B 4e-61
[SUPFAM]     kinesin motor domain homology 6e-87
[SUPFAM]     kinesin-related protein KLP4 1e-43
[SUPFAM]     kinesin-related protein nodA 1e-30
[SUPFAM]     kinesin-related protein Eg5 5e-59
[PROSITE]    ATP_GTP_A 1
[PROSITE]    KINESIN_MOTOR_DOMAIN1 1
[PFAM]       Kinesin motor domain
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY 8.57 %

```

```

SEQ      MSVTEEDLCHMMKVVRVRPENTKEKAAGFHKKVHVVDKHLVDFDPKQEEVSFFHGKKT
SEG      .....
3kar-    .....TBEEE

SEQ      NQNVIKKQNKDLKFVDAVFDETSTQCEVFEHTTKPILRSFLNGYNTVLAYGATGAGKT
SEG      .....
3kar-    EEEEETTTTTEEEEEETETTTTCHHHHHHHHHHHH-HHHGGGGCCCEEEEEECTTTTCHH

SEQ      HTMLGSADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVR
SEG      .....
3kar-    HHHHTTTT--THHHHHHHHHHHHHHHHHGGGCEEEEEEEEEETTEEEETT-TCCCEEE

SEQ      EDTQGVVVHGLTLHQPKSSEIHLHLLDNGNKNRTQHPDMNATSSRSHAVFQIYLRQOD
SEG      .....
3kar-    EETTTTEEEETTCCCEEECCGGHHHHHHHHHHHHHCCCTTTTCHHHHHHCEEEEEEEEEEE

SEQ      KTASINQNVRIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLLAGNVINALADSKR
SEG      .....
3kar-    TTTTCEE---EEEEEEECCECCCCCCC---HHHHHHHHHHHHHHHHHHHHHHHHHTTTT

SEQ      KNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKSSLK
SEG      .....
3kar-    TTTCTTTTTHHHHHHHGGGCTTTTEEEEEEECCGGHHHHHHHHHHHHH.....

SEQ      SNVLNVNHHITQYVKICNEQKAEILLKEKLYAYEQKFTNENDQAKLMISNPQKEIE
SEG      xxxxxxxx.....
3kar-    .....

SEQ      RFQEILNCLFQNRREEIRQEYLKLEMLLKENELKSFYQQOCHKQIEMMCSEDKVEKATGKR
SEG      .....
3kar-    .....

SEQ      DHRLAMLKTRRSYLEKRREELKQFDENTNWLHRVEKEMGLLSQNGHIPKELKKDLHCHH
SEG      .....
3kar-    .....

SEQ      LHLQNKDLKAQIRHMDLACLQEQHRQTEAVLNALLPTLRKQYCTLKEAGLSNAAFESD
SEG      xxx.....
3kar-    .....

SEQ      FKEIEHLVERKKVVVWADQTAEQPKQNDLPGISVLMTFPQLGPVQPIPCSSSSGGTNLVK
SEG      .....
3kar-    .....

SEQ      IPTEKRRRLKMPSPKLGQHTLKSPPSQSVQLNDSLSELQPIVYTPEDCRKAFQNPSTV
SEG      .....
3kar-    .....

SEQ      TLMKPSSFTTSFQAISNINSNCLKMLCEVAIPHNRKKECGQEDLDSTFTICEDIKSSK
SEG      .....
3kar-    .....

SEQ      CKLPQESLPNDNKDILQRLDPSSFSTKHSMPVPSMVPSYAMTTAAKRRKLTSTSTNS
SEG      .....
3kar-    .....

```

SEQ SLTADVNSGFAKRVQRQDNSSEKHLQENKPTMEHKRNICKINPSMVRKFGRNISKGNLR
 SEG xxx.....
 3kar-

Prosites for DKFZphtes3_26g22.1

PS00017 113->121 ATP_GTP_A PDOC00017
 PS00411 252->264 KINESIN_MOTOR_DOMAIN1 PDOC00343

Pfam for DKFZphtes3_26g22.1

HMM_NAME Kinesin motor domain

HMM *RCRPLNeREindgcscvVQWpPwtGyktvhngheds.....
 R+RP N +E+++G _VV + + + + + +E S

Query 17 RVRPENTKEKAAGFHKKVVHVVD-KHILVFDPKQEEVSFFHGKKTNQNV 64

HMMphksFtFDHVFwWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQ
 + F+FD VF+ ++TQ +V++ + PI+ ++++GYNCT++AYG

Query 65 IKKQNKDLKFVFDVDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGA 114

HMM TGSGKTYTMMGpggehPDHmGIIPRCCHDIFdrIdkfgekDhdFWHvKCS
 TG+GKT+TM G + D+ G+ + +++++ D + + + +S

Query 115 TGAGKTHMLG----SADEPGVMYLTMLHLYKCMDEIK-EEKIC-STAVS 158

HMM YMEIYNeeIYDLLCPnPgHmKpLnIHEHPNMGPYVqGCTefHvCSYeDac
 Y+E+YNE+I+DLL+ N ++PL+++E+ G+ V G+T+ +S E+++

Query 159 YLEVYNEQIRDLLV-N---SGPLAVREDTQKGVVHGLTLHQPKSSEEIL 204

HMM hWIWqGnknRHVAaTnMNdhSSRShtIFTIHveQrHk..qcdehvcHskM
 H+++ GNKNR+ +T MN++SSRSH++F+I ++Q K + V++ KM

Query 205 HLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQDKTASINQNVRIAKM 254

HMM NLVDLAGSERvnrTGAEGQRlKEGcNINqSLtLlGnVInaLaDggTKYmY
 +L+DLAGSER++ +GA G+R+ EG+NIN+SL++LGNVINALAD +

Query 255 SLIDLAGSERASTSGAKGTRFVEGTNINRSLALGNVINALADSK----- 299

HMM gghgHIPYRDSKLTWLLQDSLGGNCkTcMIACIWPadWNYEETLSTLRYA
 +++HIPYR SKLT+LL+DSLGGNC T MIA+++P+ + Y++T +TL+YA

Query 300 RKNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYA 349

HMM dRAKnIkNkPQINEDPcamalWRrYheQIqdmKhqL*
 +RAK+IK + N + + + +Y + + K++

Query 350 NRAKDIKSSLSNVNLVN-NHITQYVKICNEQKAEI 384

DKFZphtes3_27d1

group: metabolism

DKFZphtes3_27d1 encodes a novel 712 amino acid protein similar to ubiquitin-specific proteases (EC 3.1.2.15).

The novel protein contains both, a ubiquitin carboxyl-terminal hydrolases family 2 signature 1 and signature 2. Pfam predicts a new member of the ubiquitin carboxyl-terminal hydrolases family 2. The ubiquitin system is responsible for the turn over of proteins. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The novel protein is a new member of the ubiquitin carboxyl-terminal hydrolases family 2, represented by proteins such as yeast UBP1-16, human tre-2, human isopeptidase T and others.

The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

similarity to ubiquitin-specific proteases

complete cDNA, complete cds, 4 EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2871 bp

Poly A stretch at pos. 2836, no polyadenylation signal found

```
1 CCAAACCTGA AAGAGGTGA TTTGTAATGA TTTGCAGGGG GGCAGTGGAG
51 GCAGCGGCCA GGACTTTTCA CTTAGGAGAT CAGCATTTCG CCTGATGGAA
101 ACTGGGCGAT CCTGCAGGA CTGACCTCTG AGTTATCCAA AGGCCGACCT
151 GGGGAAAGAC TGATTTGAG GTTTAAATAG TTTTCAGATG CTTCAAGTGT
201 TGTGAACAGA GACTTGTGTT GATTATGCAT TTCTCAGCTA GACTAAATAA
251 ATGCTAGCAA TGGATACGTG CAAACATGTT GGGCAGCTGC AGCTTGCTCA
301 AGACCATTCC AGCCTCAACC CTCAGAAATG GCACTGTGTG GACTGCAACA
351 CGACCGAGTC CATTGGGCT TGCCTTAGCT GCTCCCATGT TGCCTGTGGA
401 AGATATATTG AAGAGCATGC ACTCAAGCAC TTCAAGAAA GCAGTCATCC
451 TGTTCATTG GAGGTGAATG AGATGTACGT TTTTGTGATG CTTTGTGATG
501 ATTATGTTCT GAATGATAAC GCAACTGGAG ACCTGAAGTT ACTACGACGT
551 ACATTAAGTG CCATCAAAAG TCAAAATTAT CACTGCACAA CTCGTAGTGG
601 GAGGTTTTTA CGGTCCATGG GTACAGGTGA TGATTCTTAT TTCTTACATG
651 ACGGTGCCCA ATCTCTGCTT CAAAGTGAAG ATCAACTGTA TACTGCTCTT
701 TGGCACAGGA GAAGGATACT AATGGGTAAA ATCTTCGAA CATGGTTTGA
751 ACAATCACCC ATTGGAAGAA AAAAGCAAGA AGAACCATT T CAGGAGAAAA
801 TAGTAGTAAA AAGAGAAGTA AAGAAAAAGC GGCAGGAATT GGAGTATCAA
851 GTTAAAGCAG AATTGGAAAG TATGCCTCCA AGAAGAGATT TACGTTTACA
901 AGGGCTCGCT CAGTCGACCA TAATAGAAAT AGTTTCTGTT CAGGTGCCAG
951 CACAAACGCC AGCATCACCA GCAAAAGATA AAGTACTCTC TACCTCAGAA
1001 AATGAAATAT CTCAAAAAGT CAGTGACTCC TCAGTTAAAC GAAGGCCAAT
1051 ACTAATCTCT GGTGTAACAG GATTGAGAAA TTTGGGAAAT ACTTGCTATA
1101 TGAATTTCTG TCTTCAGGTG TTGAGTCATT TACTTATTTT TCGACAATGT
1151 TTTTAAAGC TTGATCTGAA CCAATGGCTG GCTATGACTG CTAGCGAGAA
1201 GACAAGATCT TGTAAGCATC CACCAGTCAC AGATACAGTA GTATATCAAA
1251 TGAATGAATG TCAGGAAAAA GATACAGGTT TTGTTTCTC CAGACAATCA
1301 AGTCTGTCTA CAGGACTAAG TGGTGGAGCA TCAAAAGGTA GAAAGATGGA
1351 ACTTATTCAG CCAAAGGAGC CAACTTCACA GTACATTCTT CTTTGTCTATG
1401 AATTGCATAC TTTGTTCCAA GTCATGTGGT CTGGAAAGTG GGCCTTGGTC
1451 TCACCATTTG CTATGCTACA CTCAGTGTGG AGACTCATTC CTGCCTTTGG
1501 TGGTTACGCC CAACAAGACG CTCAGGAATT TCTTTGTGAA CTTTATAGATA
1551 AAATACAACG TGAATTAGAG ACAACTGGTA CCAGTTTACC AGCTCTTATC
1601 CCCACTTCTC AAAGGAAACT CATCAACAA GTTCTGAATG TTGTAATATA
1651 CATTTTTCAT GGACAACCTC TTAGTCAGGT TACATGCTCT GCATGTGACA
1701 ACAAAATCAA TACCATAGAA CCTTCTGGG ACTTGTCTAT GGAGTTTCCA
1751 GAAAGGTATC AATGCAGTGG AAAAGATATT GCTTCCCAGC CATGCTCTGGT
1801 TACTGAAATG TTGGCCAAAT TTACAGAAAC TGAAGCTTTA GAAGGAAAAA
1851 TCTAGGTATG TGACCAAGTGT AACTCAAAGC GTAGAAGGTT TTCCTCCAAA
1901 CCAGTTGTAC TCACAGAAGC CCAGAAACAA CTTATGATAT GCCACCTACC
1951 TCAGGTTCTC AGACTGCACC TCAAACGATT CAGGTGGTCA GGACGTAATA
2001 ACCGAGAGAA GATTGGTGTG CATGTTGGCT TTGAGGAAAT CTTAAACATG
2051 GAGCCCTATT GCTGCAGGGA GACCTGAAA TCCTCAGAC CAGAATGCTT
2101 TATCTATGAC TTGTCCCGCG TGGTGTATGCA CCATGGGAAA GGATTGGGCT
2151 CAGGGCACTA CACTGCCTAC TGCTATAATT CTGAAGGAGG GTTCTGGGTA
2201 CACTGCAATG ATTCCAAACT AAGCATGTGC ACTATGGATG AAGTATGCAG
2251 GGCTCAAGCT TATATCTTGT TTTATACCCA ACGAGTTACT GAGAATGGAC
```

```

2301 ATTCTAAACT TTTGCCTCCA GAGCTCCTGT TGGGGAGCCA ACATCCCAAT
2351 GAAGACGCTG ATACCTCGTC TAATGAAATC CTTAGCTGAT CCAAAGACAA
2401 TGGGGTTTTC TTCCTGTGAT TTATATATAT ACTTTTAAAG AGACTGATGT
2451 ACCATTTTAA ACTTCATTTT TTCTTGTGAA TCAGTGATATA CTACATTTAT
2501 ACATTTTATA TCTAACAATT TTTTCTTTT ACAGAGTATA AATGTATATA
2551 TCAACTGAAG GTAACACTTT TTTTCATATT TGGAGTTTAA AACTTTTGGT
2601 GTTTACCTCA GACTGATGTT ACCTCTTTTA TATTTTATG TCTTAATTGG
2651 CTCGGATGAT GAACCTGTGC AATCTTCTAC CAACAAAGTT CAAGTGGCAT
2701 CATTITATAT ACATGTATCT TTTTCAGGTA TTTTCTATAC AAATCTTAA
2751 TAGATGGAAA ATTAGACTCT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AAAAAAAAAA AAAAAAAAAA AAGGGGCGGC CGCTCTAAAA AAAAAAAAAA
2851 AAAAAAAAAA AAAAAAAAAA G

```

BLAST Results

No BLAST result

Medline entries

98072201:
Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.

98431658:
The ubiquitin system.

Peptide information for frame 2

ORF from 251 bp to 2386 bp; peptide length: 712
Category: similarity to known protein
Prosite motifs: UCH_2_1 (274-290)
UCH_2_2 (619-638)
UCH_2_2 (619-638)

```

1 MLAMDTCKHV GQLQLAQDHS SLNPQKWHCV DCNTTESIWA CLSCSHVACG
51 RYIEEHALKH FQESSHPVAL EVNEMYVFCY LCDDYVLNDN ATGDLKLLRR
101 TLSAIKSONY HCTTRSGRFL RSMGTGDDSY FLHDGAQSLL QSEDQLYDAL
151 WHRRRILMGK IFRTWFEQSP IGRKKQEEPF QEKIVVKREV KKRRQLEYEQ
201 VKAELESMPK RKSRLRLQGLA QSTIIIEIVSV QVPAQTPASP AKDKVLSTSE
251 NEISQKVSDES SVKRRPIVTP GVTGLRNLGN TCYMNVSVLQV LSHLLIFRQC
301 FLKLDLNLQWL AMTASEKTRS CKHPPVTDV VYQMNCEQEK DTGFVCSRQS
351 SLSSGLSGGA SKGRKMELIQ PKEPTSQYIS LCHLHLLFQ VMWSGKWALV
401 SPFAMLSHVW RLIPAFRGYA QQDAQEFLCE LLDKIQRELE TTGTSPLPALI
451 PTFQRKLKQ VLNVVNNIFH GQLLSQVTC ACNKNSTIE PFWDLSLEFP
501 ERYQCSGKDI ASQPCLVTEM LAKFTETEAL EGKIYVCDQC NSKRRRFSSK
551 PVVLTEAQKQ LMICHLPOVL RLHLKRRFRWS GRNNREKIGV HVGFEELNM
601 EPYCCRETLK SLRPECFIYD LSAVVMHGHK GFGSGHYTAY CYNSEGGFWV
651 HCNDSKLSMC TMDEVCKAQA YILFYTQRTV ENHGSKLLPP ELLLSQHPN
701 EDADTSSNEI LS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27d1, frame 2

PIR:S57591 hypothetical protein YMR223w - yeast (Saccharomyces cerevisiae), N = 4, Score = 218, P = 8.4e-38

SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (KIAA0055), N = 2, Score = 300, P = 9.3e-31

TREMBL:AF079565_1 gene: "Ubp41"; product: "ubiquitin-specific protease UBP41"; Mus musculus ubiquitin-specific protease UBP41 (Ubp41) mRNA, complete cds., N = 3, Score = 187, P = 8.7e-30

PIR:I58376 hypothetical protein unp - mouse, N = 3, Score = 214, P = 1.2e-28

>SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13)
 (DEUBIQUITINATING ENZYME 11) (KIAA0055).
 Length = 1,118

HSPs:

Score = 300 (45.0 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 95/301 (31%), Positives = 149/301 (49%)

Query: 381 LCHELHTLFQVMWSGKVALVSPFAMLHVSWRLLIPAFRGYAQQDAQEFLCELDDKIQREL- 439
 + E + + +W+G++ +SP ++ ++ F GY+QQD+QE L L+D + +L
 Sbjct: 826 VAEERFGIIMKALWTGQYRYISPKDFKITIGKINDQFAGYSQQDSQELLFLMDGLHEDLN 885

Query: 440 -----ETTGTSLPALIPTSQRKLIKQVLN--VVNNIFHGQLLSQVTCCLADNKSNT 488
 E L + LN ++ +F GQ S V CL C KS T
 Sbjct: 886 KADNRKRYKEENNDHLDLDFKAAEHAWQKHQKLNESIIVALFQGFQKSTVQCLTCHKKSRT 945

Query: 489 IEPFWDLSEFFPERYQCSGKDIASQCLVTEMLAKFTETELEGKIYVCDQCNSKRRRFS 548
 E F LSL +C+ +D CL + +K E + + + C C ++R
 Sbjct: 946 FEAFMYLSLPLASTSKCTLQD-----CL--RLFSK--EEKLTDNNRFYCSHCRARR----- 992

Query: 549 SKPVVLTEAQKQLMICHLPQVLRHLKRFWRSGRNNREKIGVHVGE-EILNMEPYCC-- 605
 ++ K++ I LP VL +HLKRF + GR ++K+ V F E L++ Y
 Sbjct: 993 -----DSLKKIEIWKLPVLLVHLKRFSDYDGRW-KQKLQTSVDFFLENLDSQYVIGP 1044

Query: 606 RETLKSRLPECFIYDLSAVVMHHGKGFSGHYTAYCYNSEGGFWVHCNDSKLSMCTMDEV 665
 + LK Y+L +V H+G G GHYTAYC N+ W +D ++S ++ V
 Sbjct: 1045 KNNLKK-----YNLFVSNNHYG-GLDGGHYTAYCKNAARQRFKFDDEHVS DISVSSV 1096

Query: 666 CKAQAYILFYTQ---RVTE 681
 + AYILFYT RVT+
 Sbjct: 1097 KSSAAYILFYTSLGPRVTD 1115

Score = 126 (18.9 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 41/116 (35%), Positives = 63/116 (54%)

Query: 200 QVKAEELESMPPR--KSLRLQGLAQSTIIEIVSVQVPAQTASPAPKDKVLSTSENEISQKV 257
 Q+ AE + P + +S + Q+ I+ + P TP ++K + EIS ++
 Sbjct: 701 QIPAERDREPSKLKRSYSSPDITQA--IQEEERKPTVPTVNRENKPTCYPAEIS-RL 757

Query: 258 SDSSVKKR-RPIVT---PGVTGLRNLGNTCYMNSVLQVLS---HLLIF--RQCFLKLDLQ 308
 S S ++ P+ P +TGLRNLGNTCYMNS+LQ L HL + R C+ D+N+
 Sbjct: 758 SASQIRNLNPVFGSGPALTGLRNLGNTCYMNSILQCLCNAPHLADYFNRNRYQD-DINR 816

Score = 50 (7.5 bits), Expect = 8.3e-23, Sum P(2) = 8.3e-23
 Identities = 29/106 (27%), Positives = 51/106 (48%)

Query: 173 RKKQEFPQEKIVVKREVKKRRQELEYQVKAEELESMPPRKSLRLQGLAQSTIIEIVSVQV 232
 + KQE+ +E+ +++ K R++E E + K + E+ + Q A+ + + S Q
 Sbjct: 475 KNKQEKELRERQEQEKEKLRKEEQEQAQKAKKQEA-EENEITEKQKQAKEEMKESEQA 533

Query: 233 PAQ---TPASPAKD---KVLSTSENEIS--QKVSDSSVKKRRPIVTPGV 272
 + T A K+ K S SE+E S +K + KR P TP +
 Sbjct: 534 KKEDKETSARKGKEITGVKRQSKSEHETSDAKKSVEDRGKRCP--TPEI 580

Score = 42 (6.3 bits), Expect = 5.7e-22, Sum P(2) = 5.7e-22
 Identities = 13/58 (22%), Positives = 27/58 (46%)

Query: 167 EQSPIGRKKQEFPQEKIVVKREVKKRRQELEY-QVKAEELESMPPRKSLRLQGLAQST 223
 EQ +KKQE E +++ K+ ++ E Q K E + ++ + G+ ++
 Sbjct: 498 EQEQKAKKKQEAENEITEKQKQAKEEMKESEQAQKEDKETSARKGKEITGVKRQS 555

Pedant information for DKF2phtes3_27d1, frame 2

Report for DKF2phtes3_27d1.2

(LENGTH) 712
 (MW) 81155.71
 (pI) 8.21
 (HOMOL) SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING
 ENZYME 11) (KIAA0055). 4e-32
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YMR223w] 5e-33
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YMR223w] 5e-33

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YBL067c] 3e-19
 [FUNCAT] 10.03.99 other osmosensing activities [S. cerevisiae, YDR069c] 2e-17
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR069c] 2e-17
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR069c] 2e-17
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR069c] 2e-17
 [FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YDR069c] 2e-17
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL186w] 4e-17
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YHL010c] 3e-12
 [BLOCKS] BL00970A Nuclear transition protein 2 proteins
 [BLOCKS] BL00972D
 [BLOCKS] BL00972C
 [BLOCKS] BL00972B
 [BLOCKS] BL00972A
 [EC] 3.1.2.15 Ubiquitin thiolesterase 5e-06
 [PIRKW] alternative splicing 2e-11
 [PIRKW] thiolester hydrolase 5e-06
 [PIRKW] hydrolase 1e-14
 [SUPFAM] RING finger homology 7e-11
 [SUPFAM] deubiquinating enzyme SSV7 5e-16
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 10
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] UCH_2_2_1
 [PROSITE] PKC_PHOSPHO_SITE 17
 [PROSITE] ASN_GLYCOSYLATION 4
 [PROSITE] UCH_2_1_1
 [PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
 [PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
 [KW] Alpha Beta
 [KW] LOW_COMPLEXITY 4.92 %

SEQ MLAMDTCKHVQLQLAQDHSSLNPKQWHCVDCNTTESIWACLSCSHVACGRYIEEHALKH
 SEG
 PRD cccccccccchhhhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhh

SEQ FQESSHPVALEVNMVFCYLCDDYVLNDNATGDLKLLRRTLSAISKQNYHCTTRSGRFL
 SEG
 PRD hhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhcccccccccccc

SEQ RSMGTGDDSYFLHDGAQSLLQSEDQLYTALWHRRRIILMGKIFRTWFEQSPIGRKKQEFPF
 SEG
 PRD cccccccccccccccccchhh

SEQ QEKIVVKREVKKRQLELEYQVKALESMPPRKSLRLQGLAQSTIIIEIVSVQVPAQTPASP
 SEG xxxxxxxxxxxxxxxxxxxx.....
 PRD hheehhh

SEQ AKDKVLSTSENEISQKVSVDSSVKRRPIVTPGVTLRLNLGTCYMNVLQVLSHLIFRQC
 SEG
 PRD cchhhhhhhhhhhhhhhhh

SEQ FLKLDLNQWLAMTASEKTRSCKHPPVTDTVVYQMNECQEKDTGFVCSRQSSLSGLSGGA
 SEGxxxxxxxxxxxxxxxxx.....
 PRD hhh

SEQ SKGRKMELIQPKETSQYISLCHLHTLFQVMWSGKWALVSPFAMLSHVWRLIPAFRGYA
 SEG xxxxx.....
 PRD cccccccccccccchhh

SEQ QQDAQEFLCELLDKIQRELETTGTSLPALIPTSRKLIKQVLNVVNNIFHGQLLSQVTCL
 SEG
 PRD hhh

SEQ ACDNKSNTIEPFWDLSLEFPERYQCSGKDIAEQPCLVTEMLAKFTETEALEGKIYVCDQC
 SEG
 PRD cchhhhhhhhhhhhhhhhhhh

SEQ NSKRRRFSSKPVVLTEAQQLMICHLPQVLRHLKRFWRSGRNNREKIGVHVGFEEILNM
 SEG
 PRD cccccccccchhh

SEQ EPYCCRETLSLRPECFIYDLSAVVMHKGFGSGHYTAYCYNSEGGFWVHCNDSKLSMC
 SEG
 PRD ccc

SEQ TMDEVCKAQAYILFYTRQVTENGHSLKLLPELLLSQHPNEDADTSSNEILS
 SEG
 PRD cchhh

Prosite for DKFZphtes3_27d1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00001	90->94	ASN_GLYCOSYLATION	PDOC00001
PS00001	484->488	ASN_GLYCOSYLATION	PDOC00001
PS00001	653->657	ASN_GLYCOSYLATION	PDOC00001
PS00004	545->549	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	506->509	PKC_PHOSPHO_SITE	PDOC00005
PS00005	542->545	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	580->583	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	611->614	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	223->227	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	525->529	CK2_PHOSPHO_SITE	PDOC00006
PS00006	661->665	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	193->200	TYR_PHOSPHO_SITE	PDOC00007
PS00007	192->200	TYR_PHOSPHO_SITE	PDOC00007
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	359->365	MYRISTYL	PDOC00008
PS00008	471->477	MYRISTYL	PDOC00008
PS00008	589->595	MYRISTYL	PDOC00008
PS00009	171->175	AMIDATION	PDOC00009
PS00009	362->366	AMIDATION	PDOC00009
PS00972	274->290	UCH_2_1	PDOC00750
PS00973	619->638	UCH_2_2	PDOC00750

Pfam for DKFZphtes3_27d1.2

HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*GIqNlGNTCYMNSIIQCL*		
	G++NLGNTCYMNS++Q+L		
Query	274	GLRNLGNTCYMNSVLQVL	291
HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*YdLYgVICHYGntldyGHYwaYVKNenhHRWkWYYFDDEtV*		
	YDL +V+ H+G + ++GHY+AY++N + ++W+ +D++		
Query	619	YDLSAVVMHHGKGFGSGHYTAYCYNSE--GGFWVHCNDSKL	657

DKFZphtes3_27k4

group: transmembrane protein

Summary DKFZphtes3_27k4 encodes a novel 490 amino acid protein with similarity to two hypothetical *C.elegans* proteins.

The novel protein contains 10 transmembrane regions and a leucine zipper. It is a member of the new 10 trans-membrane domain containing protein family which is specific for multicellular eukariotes.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

strong similarity to *C.elegans* K07H8.2/ZK185.2
membrane regions: 10

complete cDNA, complete cds potential start at Bp 109, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1901 bp

Poly A stretch at pos. 1866, no polyadenylation signal found

```

1 GTGATTTACC AGAAAAACCA AGAAGACAGG CACAAAAAG CAAACGGCAT
51 TTGGCAAGAT GGATTATCAA CTGCAGTACA GACTTTTAGT AATAGATCTG
101 AGCAACACAT GGAGTATCAC AGTTTCTCAG AGCAGTCTTT TCATGCCAAT
151 AATGGGCACG CATCATCAAG CTGCAGCCAA AAGTATGATG ACTATGCCAA
201 TTATAATTAC TGTGATGGAA GGGAGACTTC AGAAACCACT GCCATGTTAC
251 AAGATGAAGA TATATCTAGT GATGGTGATG AAGATGCTAT TGTAGAAGTG
301 ACCCCAAAAT TACCAAAGGA ATCCAGTGGC ATCATGGCAT TGCAAATACT
351 TGTGCCCTTT TTGCTAGCTG GTTTTGGAA CAGTTTCAGCT GGCATGGTAC
401 TGGATATAGT ACAGCACTGG GAGGTGTTCA GAAAAGTTAC AGAAGTTTTC
451 ATTTTAGTCC CTGCACTTCT TGGTCTCAA GGGAACTTGG AAATGACATT
501 GGCATCCAGA TTATCCACTG CAGTAAATAT TGGGAAGATG GATTCAACCA
551 TTGAAAAGTG GAACCTAATA ATTGGCAACT TGGCTTTAAA GCAGGTTTCA
601 GCAACAGTAG TGGGTTTCTC AGCAGCTGTG GCAGCAATTA TATTGGGCTG
651 GATTCCAGAA GGAAAATATT ACCTTGATCA TTCCATACTT CTGTGCTCTA
701 GCAGTGTGGC AACTGCCTTC ATTGCATCTC TTCTGCAGGG AATAATAATG
751 GTTGGGGTTA TCGTTGGTTC AAAGAAGACT GGTATAAATC CTGATAATGT
801 TGCTACACCC ATTGCTGCTA GTTTTGGCGA CCTTATAACT CTTGCCATAT
851 TGGCTTGGAT AAGTCAGGGC TTATACTCCT GTCTTGAGAC CTATTACTAC
901 ATTTCTCCAT TAGTTGGTGT ATTTTCTTGG GCTCTAACCC CTATTGGTAT
951 TATAATAGCT GCCAAACATC CAGCCACAAG AACAGTTCTC CACTCAGGCT
1001 GGGAGCCTGT CATAACAGCT ATGCTTATAA GTAGCATTTG GGGCCTTATT
1051 CTGGACACAA CTGTATCAGA CCCAAACTTG GTTGGGATTG TTGTTTACAC
1101 GCCAGTTATT AATGGTATTG TTGGTAATTT GGTGGCCATT CAGGCTAGCA
1151 GGATTCTTAC CTACCTCCAT TTACATAGCA TTCCAGGAGA ATTGCCTGAT
1201 GAACCCAAAG GTTGTACTA CCCATTAGA ACTTCTTTG GTCCAGGAGT
1251 AAATAATAAG TCTGCTCAAG TTCTACTGCT TTTAGTGATT CCTGGACATT
1301 TAATTTTCTT CTACACTATT CATTTGATGA AAAGTGTGCA TACTTCTTTA
1351 ACTATAATCT TCATAGTAGT GTATTTATTT GGCCTGTGTG TACAGGTATT
1401 TACCTTGCTG TGGATTGCTG ACTGGATGGT CCATCACTTC TGGAGGAAAG
1451 GAAAGGACCC GGATAGTTTC TCCATCCCTT ACCTAACAGC ATTGGGTGAT
1501 CTGCTCGGGA CAGCTCTGTT AGCCTTAAGT TTTCAATTTT TTTGGCTTAT
1551 TGGAGATCGA GATGGAGATG TTGGAGACTA ATAAATTTCT CAAACTGCCT
1601 TCAAGTTACC AAGGAAGAAA ATACACGACA ACCACTTATG GCTCTTTTTC
1651 AAAAATCTTA AATCAGTAGT TTGACTTTTG CCAGGGTAAT CTTCAAGTTGG
1701 CCCTGATTCA ATTAATGGC CTTAATTTTT TTTTAAGGAA TTTGTGTCAA
1751 AACCAGAAATG AAGAGTATTC GTGCTGCTTT TCATAGAATA AATGATAATT
1801 TGACATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAGGGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1901 G
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 109 bp to 1578 bp; peptide length: 490
Category: similarity to unknown protein

```

1 MEYHSFSEQS FHANNHGASS SCSQKYDDYA NYNYCDGRET SETTAMLQDE
51 DISSDGDDEDA IVEVTPKLPK ESSGIMALQI LVPFLLAGFG TVSAGMVLDI
101 VQHWEVFRKV TEVFILVPAL LGLKGNLEMT LASRLSTAVN IGMDSPIEK
151 WNLIIGNLAL KQVQATVVGF LAAVAAILG WIPEGKYLD HSILLCSSSV
201 ATAFIASLLQ GIIMVGIVG SKKTGINPDN VATPIAASFG DLITLAILAW
251 ISQGLYSCL EYYIISPLVG VFFLALTPIW IIIAAKHPAT RTVLHSGWEP
301 VITAMVISSI GGLILDTTVS DPNLVGIVVY TPVINGIGGN LVAIQASRIS
351 TYLHLHSIPG ELPDEPKGCY YPFRTFFGPG VNNKSAQVLL LLVIPGHLIF
401 LYTIHLMKSG HTSLTIIFIV VYLFQAVLQV FTLLWIADWM VHHFWRKGD
451 PDSFSIPYLT ALGDLTGAL LALSFLHLWL IGDROGDVGD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27k4, frame 1

TREMBL:AF036704 2 gene: "ZK185.2"; Caenorhabditis elegans cosmid
ZK185., N = 1, Score = 730, P = 3.1e-72

TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid
K07H8., N = 1, Score = 940, P = 1.7e-94

>TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8.
Length = 507

HSPs:

Score = 940 (141.0 bits), Expect = 1.7e-94, P = 1.7e-94
Identities = 204/412 (49%), Positives = 271/412 (65%)

```

Query:   68 LPKESSGIMALQILVPFLLAGFGTVSAGMVLDIVQHWEVFRKVTEVFILVPALLGLKGNL 127
          +P ESS ++ Q+L PF +AG G V AG+VL IV W +F ++ E+ ILVPALLGLKGNL
Sbjct:   82 IPAESYVLFQVLPFFAVAGLGMVFAVLVLSIVVTWPLFEEIPEILILVPALLGLKGNL 141

Query:   128 EMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGF LAAVAAILGWIPEGKY 187
          EMTLASRLST N+G MDS ++ ++I NLAL QVQATVV FLA+ A L +IP G +
Sbjct:   142 EMTLASRLSTLANLGHMDSSKQRKDVVIANLALVQVQATVVAFLASAFAAALAFIPSGDF 201

Query:   188 YLDHSILLCSSSVATAFIASLLQGIIMVGIVGSKKTGINPDNVATPIAASFGDLITLAI 247
          H L+C+SS+ATA ASL+ ++MV VIV S+K INPDNVATPIAAS GDL TL +
Sbjct:   202 DWAHGALMCASSLATACSASLVLSLLMVVIVTSRKYNINPDNVATPIAASLGDLTTLTV 261

Query:   248 LAWISQGLYSCL EYYIISPLVG VFFLALTPIWIIIAAKHPATRTVLHSGWEPVITAMVI 307
          LA+ T +++ +V V FL L P WI IA ++ T+ L+GW PVI +M+I
Sbjct:   262 LAFFGSVFLKAHNTESWLNIVIVLFLLLLPFFWIKIANENEGTQETLYNGWTPVIMSMLI 321

Query:   308 SSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPGELPDEPK 367
          SS GG IL+T V + + Y PV+NG+GGNL A+QASR+STY H G LP+E
Sbjct:   322 SSAGGFILETAVRRYH--SLSTYGPVLNGVGGNLAQVQASRLSTYFHKAGTVGVLPNEWT 379

Query:   368 GCYYPF--RTFFGPGVNNKSAQVLLLVIPGHLIFLYTIHLM---KSGHTSLTIIFIVV 421
          + R FF ++SA+VLLLLV+PGH+ F + I L K+ T +F +
Sbjct:   380 VSRFTSVQRAFFSKEWDSRSARVLLLVVPGHICFNFLIQLFTLTSKNNVTPHGPLETS 439

Query:   422 YLFQAVLQVFTLLWIADWMVHHFWRKGD PDSFSIPYLTALGDLTGALLALSF 475
          Y+ A++QV LL++ +V W+ DPD+ IPYLTALGDLTG LL + F
Sbjct:   440 YMIAAIIQVVILLFVCQLLVALLWKWKIDPDNSVIPYLTALGDLTGTLFIFV 493

```

Pendant information for DKFZphtes3_27k4, frame 1

Report for DKFZphtes3_27k4.1

[LENGTH] 490
[MW] 53266.39

SEQ	IGDRDGDVGD
SEG
PRD	eecccccccc
MEM	MM

PS000001	383->387	ASN_GLYCOSYLATION	PDOC000001
PS000004	108->112	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	65->68	PKC_PHOSPHO_SITE	PDOC000005
PS000005	221->224	PKC_PHOSPHO_SITE	PDOC000005
PS000006	5->9	CK2_PHOSPHO_SITE	PDOC000006
PS000006	54->58	CK2_PHOSPHO_SITE	PDOC000006
PS000006	146->150	CK2_PHOSPHO_SITE	PDOC000006
PS000006	238->242	CK2_PHOSPHO_SITE	PDOC000006
PS000006	257->261	CK2_PHOSPHO_SITE	PDOC000006
PS000006	296->300	CK2_PHOSPHO_SITE	PDOC000006
PS000006	318->322	CK2_PHOSPHO_SITE	PDOC000006
PS000007	25->33	TYR_PHOSPHO_SITE	PDOC000007
PS000008	90->96	MYRISTYL	PDOC000008
PS000008	122->128	MYRISTYL	PDOC000008
PS000008	216->222	MYRISTYL	PDOC000008
PS000008	220->226	MYRISTYL	PDOC000008

WO 01/12659

PCT/IB00/01496

PS00008	254->260	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00013	12->23	PROKAR_LIPOPROTEIN	PDOC00013
PS00013	248->259	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	459->481	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKF2phtes3_27k4.1)

DKFZphtes3_27014

group: testes derived

DKFZphtes3_27014 encodes a novel 358 amino acid protein with similarity to C. elegans cosmid C55A6.

The new protein contains a C3HC4 zinc finger (RING finger) signature. The ring finger structure binds two atoms of zinc, and is involved in mediating protein-protein interactions. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans C55A6.1

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: /map="6"

Insert length: 2158 bp

Poly A stretch at pos. 2137, polyadenylation signal at pos. 2120

```
1 CCGAGGCCAG AGAGAAAAGA CTGCGAGGTG GCCGCAGCTG TGGCCGGAGA
51 GCACAAAGAA TGAACCAAGCA GTGGAAGAGA AAATACTGTA AGCTGGCTGA
101 CTGCTGGTGA AGAAAATGCT TTATTTTGTG GGCAGGCATC TGTGGGATCT
151 GTAATAGAAA TATATTGGAG TAATTCAPGA TTCTGTGGTT GGCCCTTTTG
201 ACTGCTCTCT CTACAGGTTT AATTGGGCA TTTACTCATT TTCTGGGCTC
251 CAAGGACCAT GTATGTGTTG GGGATCTTCA ATATTCATGT TATTTCTCC
301 TTTGGTCTTA TATGATTGTT ACCTTTATGA AGCTTTAGTG ATTACAAAGC
351 ACTTTTGTG TCCATTTTGA CCTGAGCTTT GTAAACTCTG ATTTGCAGGA
401 TGGCTGGCTG TGGTGAAATT GATCATTCAA TAAACATGCT TCCTACAAAC
451 AGGAAAGCGA ACGAGTCCTG TTCTAATACT GCACCTTCTT TAACCGTCCC
501 TGAATGTGCC ATTTGTCTGC AAACATGTGT TCATCCAGTC AGTCTGCCCT
551 GTAAGCACGT TTTCTGCTAT CTATGTGTAA AAGGAGCTTC ATGGCTTGGA
601 AAGCGGTGTG CTCTTTGTCT ACAAGAAATT CCCGAGGATT TCCTTGACAA
651 GCCAACCTTG TTGTCAACAG AAGAAGCTCA GGCAGCAAGT AGAGGAAATG
701 GTGAATATGC ATGGTATTAT GAAGGAAGAA ATGGGTGGTG GCAGTACGAT
751 GAGCGCACTA GTAGAGAGCT GGAAGATGCT TTTCCAAAG GTAAAAAGAA
801 CACTGAAATG TTAATTGCTG GCTTTCTGTA TGTGCTGAT CTGAAAAACA
851 TGGTTCAATA TAGGAGAAAT GAACATGGAC GTCCGAGGAA GATTAAGCGA
901 GATATAATAG ATATACCAAA GAAGGGAGTA GCTGGACTTA GGCTAGACTG
951 TGATGCTAAT ACCGTAAACC TAGCAAGAGA GAGCTCTGCT GACGGAGCGG
1001 ACAGTGTATC AGCACAGAGT GGAGCTTCTG TTCAGCCCTT AGTGTCTTCT
1051 GTAAGGCCCC TAACATCAGT AGATGGTCAG TTAACAAGCC CTGCAACACC
1101 ATCCCCGTGAT GCAAGCACTT CTCTGGAAGA CTCTTTTGCT CATTTACAA
1151 TCAGTGGAGA CAACACAGCT GAAAGGAGTC ATAGGGGAGA AGGAGAAGAA
1201 GATCATGAAT CACCATCTTC AGGCAGGGTA CCAGCACCAG ACACCTCCAT
1251 TGAAGAAACT GAATCAGATG CCAGTAGTGA TAGTGAGGAT GTATCTGCAG
1301 TTGTTGACCA GCACTCCTTG ACCCAACAGA GACTTTTGGT TTCTAATGCA
1351 AACCAGACAG TACCCGATCG ATCAGATCGA TCGGSACTG ATCGATCAGT
1401 AGCAGGGGGT GGAACAGTGA GTGTCACTGT CAGATCTAGA AGCCCTGATG
1451 GACAGTGCAC AGTAACTGAA GTTTAAATAA AAATGTCTTC AGCTCCATGC
1501 TCAAGGTTGA AAGGGTTACC TGTAATTTTC TGCCACATA ACATTATACT
1551 CATCCCTAGT AGTGCAATTT GGGAGTTGGG GTGGGAAGGG GTATGGGAAG
1601 GATAGACTCA TAATTTAAAT GTCTAACATG TCTCTGTTGA GAAATTTATT
1651 TAATGTAAGG AACTTGGGTG TTAATAGTTG AGAGCTGTTT AGTAATAACC
1701 CAGTTTCTT GAGGTCTGTT TACTTTTATC TTTTAAAAA CTCTGTAGT
1751 TCTTTTGGCC AGTGTGTTG TATTATCTGT GCATTAAATG TCCTCATCTG
1801 ACTCCTGCAT TGTGCTTAT TTTTCTGCAT GGATTGGCAT AAGACCATTA
1851 CTAAATTTG GCACCTGTGA GATGTTTGAT ATTATGAACA GGAAACATAA
1901 TTTAATGTAT GAATAGATGT GAATTTGGGA TTTCAAAATA GATGAATAAC
1951 AACATATTTA TAGTAAAGTT ATTGAAATGG AAATGAAAAC AGCCAGTAAC
2001 TTATGTTTCA GAATGTTTGT AACACACTTC ATGGTGTTC CATAGCCTTT
2051 GCTGTCTAGT CTTATAGTTT GAGGTTTTTT TGGTCTGCAT TTTTCTTTT
2101 GATTACAAAA TTTATAATTT AATAAATACT AGAGTTTATC AAAAAAAG
2151 AAAAAAAG
```

BLAST Results

Entry HSG117 from database EMBL:

human STS SHGC-36270.

Score = 1148, P = 8.9e-45, identities = 240/250

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 400 bp to 1473 bp; peptide length: 358
 Category: similarity to unknown protein
 Prosite motifs: ZINC_FINGER_C3HC4 (51-61)

```

1 MAGCGEIDHS INMLPTNRKA NESCSNTAPS LTVPECAICL QTCVHPVSLP
51 CKHVFCYLCV KGASWLGKRC ALCRQEIPED FLDKPTLLSP EELKAASRGH
101 GEYAWYYEGR NGWWQYDERT SRELEDAFSK GKNTEMLIA GFLYVADLEN
151 MVQYRRNEHG RRRKIKRDII DIPKGVAGL RLDCDANTVN LARESSADGA
201 DSVSAQSGAS VQPLVSSVRP LTSVDGQLTS PATSPDAST SLEDSFAHLQ
251 LSGDNTAERS HRGEGEDHE SPSSGRVPAP DTSIEETESD ASSDSEDVSA
301 VVAQHSILTQQ RLLVSNANQT VPDRSDRSRG DRSVAGGGTV SVSVRSRRPD
351 GQCTVTEV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27o14, frame 1

TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6,
 N = 2, Score = 165, P = 4.2e-15

SWISSPROT:YW26_CAEEL HYPOTHETICAL 39.3 KD PROTEIN C02B8.6 IN CHROMOSOME
 X., N = 2, Score = 136, P = 3.1e-11

>TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6
 Length = 484

HSPs:

Score = 165 (24.8 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 42/106 (39%), Positives = 61/106 (57%)

Query: 75 QEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRN-GWWQYDERTSRELEDAFSKGGK 133
 Q +P LD ++ PEE K Y W Y G+N GWW+++ R RE+E+A++ GK
 Sbjct: 93 QNVPAALDLDA-SICDPEERK-----Y-WIYSGKNQGWWRFEPRNEREIEEAYNAGKC 142

Query: 134 NTEMLIAGFLYVADLENMVQYRRNEHGRRRKIKR---DIID-IPKKGAVAGL 180
 + E++I G YV D +QY R + R +KR D D I KG+AG+
 Sbjct: 143 HCEVVICGRPYVIDFHQFLQYPRGVPNQARHVKRVSADDFDGIGVKGLAGI 193

Score = 96 (14.4 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 19/54 (35%), Positives = 30/54 (55%)

Query: 35 ECAICLQTCVHPVSLP-CKHVFCYLCVKGASW--LGKRCALCRQEIPEDFLDKPT 86
 EC IC + P ++P C H FC++C+KG +G C +CR I + +P+
 Sbjct: 11 ECPICQCKMIVPTTIPACGHKFCFICLKGVMNDMGG-CPMCRGPIDSNIFAQPS 64

Pedant information for DKFZphtes3_27o14, frame 1

Report for DKFZphtes3_27o14.1

```

[LENGTH] 358
[MW] 38818.90
[pI] 5.17
[HOMOL] TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6 2e-12

[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR066w] 3e-04

```

```

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 4e-04
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR265w] 4e-04
[BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE] MYRISTYL 2
[PROSITE] AMIDATION 3
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] ZINC_FINGER_C3HC4 1
[PROSITE] PKC_PHOSPHO_SITE 9
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] Zinc finger, C3HC4 type (RING finger)
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 19.83 %

```

```

SEQ  MAGCGEIDHSINMLPTNRKANESCSNTAPSLTVPECAICLQTCVHPVSLPCKHVFCYLCV
SEG  .....
lrmd- .....TTTTTEETTTEETTTEEEHHHHH

SEQ  KGASWLKGRKALCRQEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRNGWWQYDERT
SEG  .....
lrmd- HHHHHHCCBTTTTTCBCGGG-CBCC.....

SEQ  SRELEDAFSKGGKNTMLIAGFLYVADLENMVQYRRNEHGRRRRIKRDIIIDIPKKGVAGL
SEG  .....XXXXXXXXXXXXXXXXX.....
lrmd- .....

SEQ  RLDCDANTVNLARESSADGADSVSAQSGASVQPLVSSVRPLTSVDGQLTSPATPSPDAST
SEG  .....XXXXXXXXXXXXX.....
lrmd- .....

SEQ  SLEDSFAHLQLSGDNTAERSHRGEGEEDHESPSSGRVPAPDTSIEETESDASSDSEDVSA
SEG  X.....XXXXXXXXXXXXXXXXXXXXX
lrmd- .....

SEQ  VVAQHSILTQQRLLVSNANQTVPDSDRSGTDRSVAGGGTVSVSVRRRPDGGQCTVTEV
SEG  xxx.....XXXXXXXXXXXXXXXXXXXXX
lrmd- .....

```

Prosite for DKF2phtes3_27ol4.1

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	318->322	ASN_GLYCOSYLATION	PDOC00001
PS00004	132->136	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	120->123	PKC_PHOSPHO_SITE	PDOC00005
PS00005	217->220	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	274->277	PKC_PHOSPHO_SITE	PDOC00005
PS00005	325->328	PKC_PHOSPHO_SITE	PDOC00005
PS00005	330->333	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	222->226	CK2_PHOSPHO_SITE	PDOC00006
PS00006	240->244	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	287->291	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00007	98->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00009	66->70	AMIDATION	PDOC00009
PS00009	130->134	AMIDATION	PDOC00009
PS00009	159->163	AMIDATION	PDOC00009
PS00518	51->61	ZINC_FINGER_C3HC4	PDOC00449

Pfam for DKFZphtes3_27ol4.1

HMM_NAME	Zinc finger, C3HC4 type (RING finger)		
HMM	*CPICFCTFQLDyPWPfdePmMlPCGHSFCypCIrrW.....CPmC*		
	C+IC	L + P++LPC+H+FCY C++	C +C
Query	36	CAIC-----LQT---CVHPVSLPCKHVFCYLCVKGASWLKGKRCALC	73

DKF2phtes3_28d14

group: testes derived

DKF2phtes3_28d14 encodes a novel 97 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1279 bp

Poly A stretch at pos. 1232, no polyadenylation signal found

```
1 GGAGCTCAGA AGTTGGGCAA AGGTCACAGC AGACTTCCTG AAAAGCAGAC
51 ACTGAGGAAC ACAGTGGAGA GCGGGAGTTC ACAGCGACGC AGCTGAGGAC
101 GACGCAGGAC CTCTCCCAAA GGTGCTGCAG CTCCAGCACC AGGGGCCAGG
151 GCTGCGGCGA CAGCAGCTCA GCAACCCCTG CTGTGCTCAA GTTCTTGGGG
201 ATTCAGAGCT AAGTTCAAAA TTAGAAACA GTGCCTTAAA GACGGGCAAG
251 AAAACCCGGT GTGGGAGTCT GCTCATCTAT GGTTTGTTCGCT TCGTCTCGCT
301 TTGATATTCT TAAATTCCTA GGTACCAATG AAAAGCCCAA GTGAACGTGG
351 CAGAGTGAGG AGGAGACAGG AGCGTGTGCA CCTTCCATCT GTGAGAGGCA
401 CACTTCAGTC TGGGTTCAAG ATGCAGAATG GTGCCTACAG CAAAAAAGAA
451 AAAAAACACC TCCTCCCTTC TTTACCATT GAATGGACAT TTTCTTACC
501 TGTGATCCCA ACAGAAACAG ATCCAGACCT ATCATGTGAA GTCCACGTTC
551 CAGGATCAGA AGTAACCAGT TTATGGACTG AGCTTACACG GGAAAGTCTA
601 CCCCCGACTC CTTCTGGATA GTAACATACA CAGCTGCATA AAAACGTCTC
651 CAAGGGGACA TACGATGCAT TTGCTTGGTG TCCAGCCCAA GCTCCCCACC
701 GGCAGACTCA CTGTTCCTTA GAGCTCGAGA GCTCGTCTCC TATCAATCAG
751 AGAACCCCAT CAGCTGTGAC CAACAGAGCT GGAGCCCTCT GTGGAGGGAG
801 CTGACCCAC ACACAGGACA GAGCAGAATC CTGATTATTT TACAACTGCG
851 AAACCTTCTG AGTAAGAAGA CAAAAATATA CATTCCAAGG TATCTGTAAA
901 GTGCTTGGAA GATGCAGACA GCTGCACCGA GGGGCTCTGA TCCATCCACA
951 CGCTGCGCTT TGCTGCGGTC ACACACAGG TCTCAGTCAC GTGATGGTTT
1001 TGCTTTTATT TCTTAAACGG CTGAGTGATA ATCCAGCTAG TGTGCAGTCA
1051 TTTCATACCT TTCAATGGGC GTCACCGCAG TGACGCTGCC CCAGCCCAT
1101 GCTGAGGGCC GACACAATTC ACGGAACAGA TTCATCATAT TTGGTCTTTA
1151 TGTAAATAAT AAATGTTTAA AAATTGCCCTA AATATAAAAA AAAAAAAAAA
1201 AAAAAAAAAA AAAAAAAAAA AAAGGGCGGC CGAAAAAAAA AAAAAAAAAA
1251 AAAAAAAAAA AAAAAAAAAA GGGCGGCCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 618 bp; peptide length: 97

Category: putative protein

```
1 MKKPSERGRV RRRQERVHLP SVRGTLQSGF KMQNGAYSKK KNNTLLPSLP
51 FEWTFSLPVI PTETDPLSC EVHVPGEVLT SLWTELTRES LPPTPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_28d14, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_28d14, frame 1

Report for DKFZphtes3_28d14.1

{LENGTH} 97
{MW} 10945.56
{pI} 9.80
{PROSITE} MYRISTYL 2
{PROSITE} CAMP_PHOSPHO_SITE 2
{PROSITE} CK2_PHOSPHO_SITE 2
{PROSITE} PKC_PHOSPHO_SITE 3
{KW} All_Alpha
{KW} LOW_COMPLEXITY 12.37 %

SEQ MKKPSEGRVRRRQERVHLPSVRGTLQSGFKMQNGAYSKKKKNTLLPSLPFEWTFSLPVI
SEGXXXXXXXXXX.....
PRD cccccchhhhhhhhhcc

SEQ PTETDPDLSCEVHVPGEVTSWTELTRESLPPTPSG
SEG
PRD cccccccceeeccccchhhhhhhhhccccccc

Prosites for DKFZphtes3_28d14.1

PS00004	2->6	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	41->45	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_28d14.1)

DKF2phtes3_2a11

group: testes derived

DKF2phtes3_2a11 encodes a novel 1048 amino acid protein with very weak similarity to mucins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to mucin

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4082 bp

Poly A stretch at pos. 4060, polyadenylation signal at pos. 4034

```
1 GAGGACTGCG AGCACAGCGG CGGCCGGGTG GCGGGGGTGA GTGGGGCCAG
51 CGGGGCTGGA CAGCAGCGGG CCGCGGGGCG CGCCGCCGCG ATCCCTCCCC
101 GCGCCCGCGC AGCACATCGC CGCCGCCGAG ATGGGCCCTC CGCGGCACCC
151 CCAGGCGCGG GAGATAGAAG CGGCGGTGCG GGGCGGCGGG CGCGGGCTAC
201 AGGTGGAAAT GAGTTCTCAA CAGTTTCCTC GGTAGGAGC CCCTTCTACC
251 GGGCTGAGCC AGGCCCTTTC TCAGATTGCA AACAGTGGTT CTGCTGGATT
301 GATAAACCCA GCTGCTACAG TCAATGATGA ATCTGGTCCA GATTCTGAAG
351 TCAGTGCCAG GGAGCACATG AGTTCCAGCA GTCCTCTCCA GTCCCGGGAG
401 GAGAAGCAAG AGCCTGTGTG GGTAAAGGCC TATCCACAGG TGCAGATGTT
451 GTCGACACAC CATGCTGTCT CATCAGCCAC ACCTGTTGCA GTGACAGCCC
501 CGCCAGCACA CCTGACGCCA GCAGTGCCAC TTTCATTTTC GGAGGGACTT
551 ATGAAGCCCG CCGGAAGGCC CACCATGCCT AGCCGTCCCA TTGCTCCTGC
601 TCCACCTTCT ACCCTGTGAC TTCCCCCCAA GGTCCAGGGC CAGGTTACCG
651 TTACCATGGA GAGTAGCATC CCTCAAGCTT CAGCCATTCC TGTGGCAACA
701 ATCAGTGGAC AACAGGGCCA TCCAGTAAC CTGCATCACA TCATGACTAC
751 AAATGTGCAA ATGTCTATCA TCCCGAGCAA TGCTCCTGGG CCCCCCTCTC
801 ACATTGGAGC TTCTCATTTA CCTCGAGGTG CAGCTGCTGC TGCTGTGATG
851 TCCAGTTCTA AAGTAACCAC AGTCTTGAGG CCGACCTCAC AGCTGCCAAA
901 TGCTGCTACT GCTCAGCCAG CAGTACAGCA CATCATTCAC CAACCAATCC
951 AGTCTCGGCC ACCTGTGACC ACCTCCAATG CCATCCCTCC TGCTGTGGTA
1001 GCAACTGTCT CAGCCACCAG AGCTCAGTCT CCAGTCATCA CTACGACAGC
1051 GCGCGATGCT ACTGATTGAG CACTTAGTAG GCCAACCTTG TCATCCAGC
1101 ATCTCTCATC TGCAGCAATC AGTATTGAGC GTCCTGCCCA GTCACGAGAT
1151 GTCACAACAA GAATCACTAC ACCATCTCAC CCTGCATTAG GGACGCCAAA
1201 ACAGCAGCTT CATACAATGG CTCAGAAAAC AATCTTCAGT ACTGGCAGCG
1251 CAGTGGCTGC AGCCACAGTA GCACCTATTT TGGCAACCAA CACCATTCCT
1301 TCAGCGACCA CAGCTGGATC TGTGTACAC ACCCAAGCTC CCACAAGTAC
1351 CATTTGTACC ATGACAGTAC CCTCCCATTC CTCCCATGCT ACTGCTGTGA
1401 CCACCTCAAA CATCCAGTCC GCCAAGGTGG TGCCCCAGCA GATCAGGCAC
1451 ACTTCTCCTC GGATCCAGCC AGACTACCTT GCCGAGAGGA GTAGCCTGAT
1501 TCCCATCTCC GGACATCGGG CCTCTCCCAA TCCTGTGGCC ATGGAAACCC
1551 GAAGTGACAA CAGACCGTCT GTTCCCGTTC AGTTCCAATA TTTTGTGCCA
1601 ACTTACCCCC CTCTGCATA CCCACTGGCG GCACATACCT ACACCCCAAT
1651 CACCAAGTCC GTGTCCACTA TCCGACAGTA TCCAGTTTCA GCTCAGGCTC
1701 CAAACTCTGC CATCACAGCT CAGACTGGTG TTGGGGTAGC GTCTACCGTC
1751 CACCTAAACC CCATGCAGTT GATGACAGTG GATGCATCGC ATGCTCGACA
1801 TATTCAAGGG ATCCAGCCAG CACCCATCAG TACCCAGGGT ATCCAGCCGG
1851 CCCCATTGG GACCCAGGG ATACAGCCTG CACCATTGG CACACAGGGA
1901 ATTCACTCAG CAACCCCAAT CAACACACAA GGGCTTCAGC CTGCACCTAT
1951 GGGTACTCAG CAGCCTCAGC CTGAAGGAAA GACTTCAGCA GTGGTGTGG
2001 CAGATGGAGC CACAATTGTG GCCAACCCTA TTAGCAATCC ATTCAGTGGT
2051 GCTCCAGCAG CAACAACCGT GGTGCAGACC CACAGCCAGA GTGCTAGCAC
2101 CAACGCTCCC GCCCAGGGCT CATCGCCACG GCCAAGCATA CTCCGGAAGA
2151 AACCTGCCAC AGATGGTGCC AAACCCAAAG CTGAAATCCA CGTGTCTATG
2201 GCCACTCCGG TCACTGTGTC CATGGAGACT GTATCCAATC AAAATAATGA
2251 TCAGCCTACC ATTGCCGTCC CTCCAAGTCC CCAGCAGCCC CCACCGACCA
2301 TTCCAATAT GATTGCAGCA GCCAGTCCCC CGTCACAACC AGCCGTGGCC
2351 CTTTCAACCA TTCCTGGAGC GGTCCCCATC ACTCCACCCA TCACCAACAT
2401 TGCAAGTGCA CCACCTCCAT CAGTCACTGT GGGTGGCAGT CTTTCTCCG
2451 TCTTGGGCCC TCCCGTTCCT GAAATTAAG TGAAGAAGA AGTAGAACCA
2501 ATGGATATCA TGAGGCCAGT TTCTGCAGTT CCTCACTGG CTACCAACAC
2551 TGTGTCTCCA TCTCTGCAT TGCTGGCAAA CAACTTGTC ATGCCTACAA
2601 GTGACCTACC ACCTGGTGCC TCCCAAGGA AAAAGCCTCG AAAGCAACAG
2651 CATGTGATCT CAACAGAAGA AGGTGACATG ATGGAGACAA ACAGCACTGA
2701 TGATGAGAAG TCCACTGCCA AGAGTCTTCT GGTGAAGGCT GAGAAGCGCA
```

```

2751 AGTCTCCTCC CAAGGAGTAT ATTGATGAGG AAGGTGTGAG ATATGTCCCA
2801 GTGCGTCCAA GACCCCCCAT TACTTTGCTT CGTCACATC GGAACCCCTG
2851 GAAAGCTGCT TACCACCACT TTCAGAGGTA CAGTGACGTC CGGGTCAAAG
2901 AGGAGAAGAA AGCTATGCTG CAGGAAATAG CTAATCAGAA AGGAGTATCC
2951 TGTCTGTGCTC AAGGCTGGAA AGTCCACCTC TGTGCTGCCC AGTTACTACA
3001 GCTGACGAAT CTAGAACATG ATGTCTATGA AAGACTTACT AACCTGCAGG
3051 AAGGGATTAT CCCAAAGAAA AAAGCAGCAA CAGATGATGA TCTCCACCGA
3101 ATAAACGAAC TGATACAGGG AAATATGCAG AGGTGTAAAC TTGTGATGGA
3151 TCAATCAGT GAAGCCAGAG ACTCCATGCT TAAGGTTTAA GATCATAAAG
3201 ACCGTGTCTT GAAGCTGCTT AACAAAGAAC GGACTGTCAA AAAAGTGTCC
3251 AAATTGAAGC GAAAGGAAAA AGTCTAGACC CAGAACATC AGGAGATTGG
3301 AAGCAAATTT ATGAAGAATG ATGGTGGGGG TGGGGGGAGG GTTTTGGTTT
3351 TTTCCAAAGT GGAACATTGA AATAAAGGAA GTGTTCTTAA GTTCCCCTGT
3401 GAAAGCAGAG GAACCCATGA CATCCAAGGG CGTGAAAGGA TCAGAGCTGA
3451 CTGGACATAG TGAGCTGCCT TCTTGCCTTC GGGTGCACCC CTGTTAAACC
3501 TGATCTGTGT CATAAGTGAC TCCGGATGCA TCAGTGTCCA CCAGTTGGAA
3551 GCAATGACAA GGATGGCTGG CTGGTGT TTTT TCAGCCTTCC GGTTTATAGA
3601 CTGTATTAT CTAGTGGATT CCTGCAGGCC CCATACTGAG CCTGGACTGA
3651 AAGTATCCAC TCGGACCATC TGTATCTCT CTACACTGAA AATAAACCTT
3701 CTTCACCCA CCCCATTCGG TTCTTCTGCC TGACCTTCAA ATGCCCATGT
3751 TGCCCTTTTA CAGCAGTGCC ACGGCACCAA GCGAGCTGCC ACATCTCACA
3801 CTCTAAAGGG TTTGAACTAT TAGTCTTGT CATTTTTAA AAAAAACCAT
3851 TCCCAGTGA AATTGTTATA TCGTCTGTCT TCGTGTGTCT AGAACTGGGT
3901 TTTTGTGGAG GTTCAGAGCA GGCAACACCA TAAGTTGCTC TCAGATCCTT
3951 GTTCTGAAGT ACATTCTTGG TTATCTGTAC TTCTGTAGCT GGTGTGATGC
4001 TGTTAATTGT ATGTACCACA CATCTCCAGA CGTTAATAAA GGACTCAAAG
4051 AGGTTTTTGT AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 131 bp to 3274 bp; peptide length: 1048
 Category: similarity to known protein

```

1 MGPPRHPQAG EIEAGGAGGG RRLQVEMSSQ QFPRLGAPST GLSQAPSQIA
51 NSGSAGLINP AATVNDESGR DSEVSAREHM SSSSSLSQRE EKQEPVVVRP
101 YPQVQMLSTH HAVASATPVA VTAPPALHTP AVPLSFSEGL MKPPFKPTMP
151 SRPIAPAPPS TSLSPKVPVG QVTVTMESSI PQASAIPVAT ISGQGHPSN
201 LHHIMTTNVQ MSIIRSNAPG PPLHIGASHL PRGAAAAAVM SSSKVTTLVR
251 PTSQLPNAAT AQPAPQHIIH QPIQSRPPVT TSNAIPPAVV ATVSATRAQS
301 PVITTTAAHA TDSALSRPTL SIQHPPSAAI SIORPAQSRD VTTRITLPSH
351 PALGTPKQQL HTMAQKTIFS TCTPVAAATV APILATNTIP SATTAGSVSH
401 TQAPTSTIVT MTVPSSHSHA TAVTTSNIPV AKVVPPQIITH TSPRIQPDYP
451 AERSSLIPIS GHRASPNPVA METRSDNRPS VPVQFYFLP TYPPSAYPLA
501 AHTYTPITSS VSTIRQYPVS AQAPNSAITA QTGVGVASTV HLNPMQLMTV
551 DASHARHIQI IQPAPISTQG IQPAPIGTPG IQPAPLGTQG IHSATPINTQ
601 GLQAPRMGTQ QPQPEGKTSV VVLADGATIV ANPISNPFSA APAATTVVQT
651 HSQSASTNAP AQGSSPRPSI LRKKPATDGA KPKSEIHVSM ATPVTVMET
701 VSNQNNDOPT IAVPPTAQPF PPTIPTMIAA ASPPSQPAVA LSTIPGAVPI
751 TPTITTIAAA PPPSVTVGGS LSSVLGPPVP EIKVKEEVEP MDIMRPVSAV
801 PPLATNTVSP SLALLANNLS MPTSDLPPGA SPRKKPRKQQ HVISTEEGDM
851 METNSTDDEK STAKSLLVKA EKRKSPKKEY IDEEGVRYVP VRPRPITLL
901 RHYRNPWKAA YHHFQRYSDV RVKEKKAML QEIANQKGVs CRAQGWKVHL
951 CAAQLQLTN LEHDVYERLT NLQEGIIKK KAATDDDLHR INELIQGNMQ
1001 RCKLVMDQIS EARDSMLKVL DHKDRVLKLL NKNGTVKKVS KLKRKEKV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2a11, frame 2

SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2)., N = 1,
 Score = 334, P = 2.4e-25

>SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
Length = 5,179

HSPs:

Score = 334 (50.1 bits), Expect = 2.4e-25, P = 2.4e-25
Identities = 184/770 (23%), Positives = 263/770 (34%)

Query:	96	VVVRPPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPKPTMPSRPI	154
		V P P T + + T V T P TP + + P P P T P	
Sbjct:	3471	VTPTPTPTGTGTPTTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTGTPT	3530
Query:	155	A-PAPPSTLSLPPKVP-QQVTVTMESSIPOASAIIPVATISGQQGHPSNLHHIMTTNVQMS	212
		P +T P P G T T + P T +G Q P+ TT V +	
Sbjct:	3531	TTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTGT-PTTPIITTTTIVTPT	3589
Query:	213	IIRSNAAGP---PLHIGASHLPRGAAAAA-VMSSSKVTVTLRPTSQLPNAATAQPAVQHI	268
		+ P P+ + P + + + +T T +PT T P I	
Sbjct:	3590	PTPTGTQTPTTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTQTPTTTPI	3649
Query:	269	IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSEVITTTAAHATDSALSRLSLIQHPSPA	328
		+ P T P T + T + +P T T + +T+ P	
Sbjct:	3650	TTTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPT	3706
Query:	329	AISIQRPQASRDVTTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPTVAAT--VAPILA	385
		Q P + TT T P+ GT + T + T TP T PI	
Sbjct:	3707	PTGTQTPTTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT	3766
Query:	386	TNTI-PSATTAGSVSHQTAPTSTIVTMT-VPSHSHATAVTTSNIPKVVVPPQIHTSP	443
		T T+ P+ T G+ + T P T+ T T P+ + T T V P T T	
Sbjct:	3767	TTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTGT	3825
Query:	444	RIQPDYPAERSSSLIPISGHRASPNPVMATESDNRNRPVQFQYFL-PTYPPSAYPLAAH	502
		+ P ++ + +P P +T + +T + PT P+	
Sbjct:	3826	QTPTTTPIITTTTIVT----PTPTPTGTGTPT---TTPITTTTIVTPTPTPTG--TQTP	3874
Query:	503	TYTPTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARIHQ	560
		T TPIT++ + T P Q P+ IT T V T Q T	
Sbjct:	3875	TTTPIITTTTIVTPTPTGTGTPTTTPIITTTTIVTPTPTP--TGTQTPTTTPIITTTTIVT	3932
Query:	561	IQPAPISTQGIQAPAGTPTGI---QPAPLGTQGIHSATPINTQGL---QFAPMGTTQQPQ-	613
		P P TQ PI T P P GTQ + TPI T P P GTQ P	
Sbjct:	3933	PTPTPTGTQTPTTTPIITTTTIVTPTPTGTGTQ-TPTTPIITTTTIVTPTPTPTGTQTPTT	3991
Query:	614	-PEGKTSAVVLADGATIANVPISNPFSAAPAA-TVVQTHSQSASTNAPAQSSPRPSIL	671
		P T+ V T P P+ + T T T+Q+ T ++ P+	
Sbjct:	3992	TPITTTTIVTPTPTPTGTGTPTTTPIITTTTIVTPTPTPTGTGTPTTTPIITTTTIVTPTPT	4051
Query:	672	RKKPATDGAKPKSEIHVSMAPTVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI	728
		T P + TP +T + T P PT Q P T P	
Sbjct:	4052	PTGTQTPTTTPIITTTTIVTPTPTPTGTGTPTTTPIITTTTIVTPTPTGTGTPTTTPIIT	4111
Query:	729	AAASPPSQPAVALSTIPGAVPTPTTITIAAAPPS-----VTVGSLSSVLGP-FVPEI	782
		P+ T P PIT TT+ P P+ T + + + P P P	
Sbjct:	4112	TTTIVTPTPTPTGTGT-PTTPIITTT-TTVPTPTPTGTGTPTTTPIITTTTIVTPTPTPTG	4169
Query:	783	KVKEEVEPMDIMRPVSAVP-PLATNTVSPSLALLANLMSPTSDLPPGASPRKKRPQQH	841
		P+ P+ P T T P+ A + TS+ P P S + R	
Sbjct:	4170	TQTPTTTPIITTTTIVTPTPTPTGTGTGPPTHSTAPIAEELTSNPPPESSPTQTSRSTSS	4229
Query:	842	VISTEEGDMMET 853	
		+ TE ++ T	
Sbjct:	4230	PL-TESTLLST 4240	

Score = 328 (49.2 bits), Expect = 1.0e-24, P = 1.0e-24
Identities = 180/745 (24%), Positives = 254/745 (34%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P T P
Sbjct: 3540 VTPTPTPTGTQTPTTTTPTTTTTVPTPTPTGTQTPTTTTPTTTTTVPTPTPTGTQTPT 3599

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3600 TTPITTTTPTPTPTGTPTTTPIITTTTPTPTPTGTPT-PTTPTITTTTPTPT 3658

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3659 PTPGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTP 3718

Query: 269 IHQPIQSRPFVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3719 TTTTPTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPT 3775

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3776 PTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITT 3835

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTTSP 443
T + P + T G + + T P +T T+T P+ + T TT V P T T
Sbjct: 3836 TTTVPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTPTPTGT 3894

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3895 QTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTPTG--TQTP 3943

Query: 503 TYTPTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 3944 TTTPTTTTPTPTPTGTQTPTTTPITTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 4001

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQFPQ- 613
P P TQ P I T P P GTQ + TPI T P P GTQ P
Sbjct: 4002 PTPPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTT 4060

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSPPRSIL 671
P T + V T P + P + T T +Q+ +T ++ P+
Sbjct: 4061 TPTTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTPTPT 4120

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4121 PTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITT 4180

Query: 729 AAASPPSQPAVALSTIPGAVPITPITTIAAA-PPPSVTVGSLSSVLGPPVPEIKVKEE 787
P+ T P T PI + + PPP + + S P +
Sbjct: 4181 TTTVPTPTPTGTQTPTTHTSTAPIAELTTSNPPPESSPTQTSRSTSSPLTESTTLLST 4240

Query: 788 VEPMDIMRPVSAVPLATNTVSPSLALLANNLSMP--TSDLPFGASPR 833
+ P M S PP +T T +P+ + LS P T+ PPG R
Sbjct: 4241 LPPAIEM--TSTAPP-STPT-APTTTSGGHTLSPPTTTSPPGTPT 4284

Score = 325 (48.8 bits), Expect = 2.2e-24, P = 2.2e-24
Identities = 186/782 (23%), Positives = 261/782 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPP AHL-TPAVPLSFSEGLMKPPPKPTMPSRFI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3494 VTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPT 3553

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3554 TTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPT-PTTPTITTTTPTPT 3612

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3613 PTPGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTP 3672

Query: 269 IHQPIQSRPFVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3673 TTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTPTPT 3729

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3730 PTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITT 3789

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTTSP 443
T + P + T G + + T P +T T+T P+ + T TT V P T T
Sbjct: 3790 TTTVPTPTPTGTQTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTPTPTGT 3848

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3849 QTPTTTPITTTTPTPTPTGTQTPTTTPITTTTPTPTPTGTPTG--TQTP 3897

Query: 503 TYTPTSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560

T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 3898 TTTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPT--TGTQTPTTPIITTTTVP 3955
 Query: 561 IQPAPISTQGIQAPAPIGTPI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 3956 PTPTPTGTQTPTTTPIITTTTVPPTPTGTGTQ-TPTTPIITTTTVPPTPTGTGTPTT 4014
 Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 4015 TPITTTTVPPTPTGTGTPTTTPIITTTTVPPTPTGTGTPTTTPIITTTTVPPTPT 4074
 Query: 672 RKKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P
 Sbjct: 4075 PTGTPTTTTPIITTTTVPPTPTGTGTPTTTPIITTTTVPPTPTGTGTPTTTPIIT 4134
 Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P I V
 Sbjct: 4135 TTTVTPPTPTGTQT-PTTTPIT---TTTTVTPPTPT--GTQT---PTTTPITTTTVP 4184
 Query: 789 EPMDIRPVSAPVPLATNTVSPSLALLANLSMPTSDLPSPGASPRKKPRKQHVISTEEG 848
 P PP T+T +P L +N P S P + P + + +
 Sbjct: 4185 TPTPTPTGTGTGPPHTST-APIAELTTSN-PPESSTPQTSRSTSSPLTESTTLLSTLP 4242
 Query: 849 DMMETNSTDDEKSTAKSLLVKAERKSPP 877
 +E ST + SPP
 Sbjct: 4243 PAIEMTSTAPPSTPTAPTTTSGGHTLSPP 4271
 Score = 324 (48.6 bits), Expect = 2.8e-24, P = 2.8e-24
 Identities = 170/717 (23%), Positives = 248/717 (34%)
 Query: 95 PVVVRFPYQVQMLSTHVASATP--VAVTAPPAHLTFAVPLSFSEGLMKPPKPTMPSR 152
 P P P +T + +P T PP TP+ P++ + + P P+ P
 Sbjct: 1401 PPTTTPSPPTTTTTLPTTTTPSPPTTTTTPPTTTPSPITTTTTPPL-PTTTPSPPT 1459
 Query: 153 PIAPAPFSTLSLPPKVPQVTVTMESSIPQASAIPTATISGQGHPSNLHHIMTINVQMS 212
 PP+T PP T S + P T + P I +
 Sbjct: 1460 TTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPS---PMTTPTITPPASTTT 1516
 Query: 213 IIRSNAPGPPPLHIGASHLPRGAAAAVMSSSKVTTLRPTSQ--LPNAATAQPAVQHIIH 270
 + + P PP + P S T + PTS LP T P
 Sbjct: 1517 LPPTTTPSPPTTTTTPPP-----TTTTPSPTTTPITPTSTTTLPTTTPSPPTTTT 1571
 Query: 271 QPIQSRP-PVTSNAIPPAVVATVSA-TRAQSPVITTTAAHATDSALSRTLSIQHPPSA 328
 P + P P TT+ PP + T T SP TTT + S PT + PP++
 Sbjct: 1572 PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTP 1631
 Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIESTGTVAATVAPILATNT 388
 ++ T T P P TP T I +T TP T + + T
 Sbjct: 1632 TTTLPPTTTPSPPTTTTTP--PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTP 1689
 Query: 389 IPSATTAGSVSHTOAPTSTIVTMTVPSHSHATAV-TTSNIPVAKVVPQOITHTSPRIQP 447
 P TT + S T P+S I T T PS ++ + TT P P T T + P
 Sbjct: 1690 SPPTTMTTTPSPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1749
 Query: 448 DYPAERSSLIPIGHRASPNFVAMETRSDNRPSVPV-QFYFLPTYPPSAY-P-----LA 500
 + + P+ P T + P VP+ + +L + P+ + P L
 Sbjct: 1750 TTTSSPLTTPLPPSITPPTFPSTTTTTPCPLCNWGLDVGKPNFHKPGGDTCLI 1809
 Query: 501 AHYTFITSSVSTIR--QYP-VSAQAPNSAITAQTVGV-VASTVHLNPMQLMTVDASHAR 556
 P ++ + R YP V + VG + P ++ + A
 Sbjct: 1810 GDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPM-AFCLN 1868
 Query: 557 HIQIQPAPISTQGIQAPAPIGTPIQ-PAPLGTQGIHSATPINTQGLQAPAPMGTOQPQ-- 613
 + +Q TQ P + T + P P T I + T + P P GTQ P
 Sbjct: 1869 YEINVQCCCEVTQ---PTTMTTTTENPTPTTPTTPIITTTTVPPT--PTPTGTQTPTT 1922
 Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSILR 672
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 1923 PIITTTTVPPTPTGTGTPTTTPIITTTTVPPTPTGTGTPTTTPIITTTTVPPTPTPT 1982
 Query: 673 KKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMIA 729
 T P + TP +T + T P PT Q P T P
 Sbjct: 1983 TGTQTPTTTTPIITTTTVPPTPTGTGTPTTTPIITTTTVPPTPTGTGTPTTTPIIT 2042
 Query: 730 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEVE 789
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 2043 TTVTPTPTGTQT-PTTTPIT---TTTTVTPPTPT--GTQTPTTPIITTTTVPPTPT 2096
 Query: 790 PMDIRPVSAPVPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 2097 PTGTQTPTTT-PITTTTVPPTPT 2117

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2068 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 2127

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQOGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2128 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 2186

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2187 PTPTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 2246

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPFSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2247 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT 2303

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2304 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2363

Query: 386 TNTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTTSP 443
T T+ P+ T G+ + T P +T T+ T TT V P T T
Sbjct: 2364 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGT 2422

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVAMETRSNRPVSVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + +P+ + PT P+
Sbjct: 2423 QTPTTTPIITTTTPTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGTGT--TQTP 2471

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2472 TTTTPTTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTPTPT--TGQTPTTTPIITTTTPT 2529

Query: 561 IQPAPISQGIQAPIGTPIG--QPAPLGTQGIHSATPINTQGL--QPAPMGTTQGPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2530 PTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-TPTTTPIITTTTPTPTPTGTGTPTT 2588

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2589 TPITTTTGTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT 2648

Query: 672 RKKPATDGAAPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 2649 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGTGTPTTTPIIT 2708

Query: 729 AAASPPSQPAVALSTIPGAVPITPPIITIAAAPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 2709 TTTTPTPTPTGTGT-PTTTPIT---TTTTPTPTPTPT--GTGTPTTTPIITTTTPTPTPT 2762

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 2763 TPTGTGTPTTT-PITTTTPTPTPT 2784

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2206 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 2265

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQOGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2266 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPTPT 2324

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2325 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 2384

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPFSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2385 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGT 2441

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2442 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2501

Query: 386 TNTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTTSP 443
T T+ P+ T G+ + T P +T T+ T P+ + T TT V P T T

Sbjct: 2502 TTTVTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPTGTGT 2560

Query: 444 RIQPDYPAERSSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+

Sbjct: 2561 QTPTTTPITTTTPTV-----PTPTPTGTGTPT-----TTPITTTTPTPTPTPTGT--TQTP 2609

Query: 503 TYTPITSSVS-TIRQYFVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2610 TTTPIITTTTPTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT--TGTQTPTTPTITTTTPT 2667

Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2668 PTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-TPTTTPITTTTPTPTPTGTGTPTT 2726

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2727 TPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPT 2786

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P

Sbjct: 2787 PTGTQTPTTPTITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTPTITT 2846

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2847 TTTVTPTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 2900

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P + P T V+P+

Sbjct: 2901 TPTGTQTPTTT-PITTTTPTPT 2922

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P

Sbjct: 2321 VTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPT 2380

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPOASAI PVATISGQGHPSNLHHIMTTNVOMS 212
P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 2381 TTPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-PTTTPITTTTPTPTPT 2439

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLENAATAQPAVQHI 268
+ P P + P +++ +TT T T P I

Sbjct: 2440 PTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTPI 2499

Query: 269 IHQPIQSRPFVTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T+ P

Sbjct: 2500 TTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT 2556

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI

Sbjct: 2557 PTGTQTPTTPTITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTPTITT 2616

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T

Sbjct: 2617 TTTVTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT 2675

Query: 444 RIQPDYPAERSSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+

Sbjct: 2676 QTPTTTPITTTTPTV-----PTPTPTGTGTPT-----TTPITTTTPTPTPTPTGT--TQTP 2724

Query: 503 TYTPITSSVS-TIRQYFVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2725 TTTPIITTTTPTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT--TGTQTPTTPTITTTTPT 2782

Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2783 PTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-TPTTTPITTTTPTPTPTGTGTPTT 2841

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2842 TPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPT 2901

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P

Sbjct: 2902 PTGTQTPTTPTITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTPTITT 2961

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2962 TTTVTPTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 3015

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811
P + P T TV+P+
Sbjct: 3016 TPTGTQTPTTT-PITTTTIVTPT 3037

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2390 VTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP 2449

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2450 TTPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP-TTTTPIITTTTIVTPT 2508

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2509 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2568

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLSLIQHPPSA 328
+ P T P T +T +P T T T + T++ P
Sbjct: 2569 TTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2625

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2626 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2685

Query: 386 TNII-PSATTAGSVSHTOAPTSTIVTMT-VPSHSSHATAVTTNSIPVAKVVPQQITHSTP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2686 TTTVTPPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTP-TPTGT 2744

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2745 QTPTTPIITTTTIVT---PTPTGTQTP---TTPITTTTIVTPTPTGT--TQTP 2793

Query: 503 TYTPIITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2794 TTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPT--TGTQTPTTTPIITTTTIVT 2851

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAFMGTQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2852 PTPTPTGTQTPTTTPIITTTTIVTPTPTGTQ-TPTTPIITTTTIVTPTPTGTQTPTT 2910

Query: 614 -PEGKTSAAVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2911 TPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPT 2970

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 2971 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 3030

Query: 729 AAASPPSQPAVALSTIPGAVPITPITIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 3031 TTTVTPPTPTGTQTP-TTTPIT---TTTIVTPTPT--GTQTPTTTPIITTTTIVTPT 3084

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811
P + P T TV+P+
Sbjct: 3085 TPTGTQTPTTT-PITTTTIVTPT 3106

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2459 VTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP 2518

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2519 TTPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP-TTTTPIITTTTIVTPT 2577

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2578 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2637

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLSLIQHPPSA 328
+ P T P T +T +P T T T + T++ P
Sbjct: 2638 TTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2694

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2695 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2754

Query: 386 TNNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2755 TTTVTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTGT 2813

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2814 QTPTTTTPIITTTTPT-----PTPTPTGTQTPT-----TPTITTTTPTPTPTPTG--TQTP 2862

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2863 TTTPIITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTG--TGTQTPTTTPIITTTTPT 2920

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2921 PTPTPTGTQTPTTTTPIITTTTPTPTPTGTQ-TPTTPTITTTTPTPTPTGTQTPTT 2979

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2980 TPIITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPT 3039

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 3040 PTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIIT 3099

Query: 729 AAASPPSQPAVALSTIPGAVPITPPIITIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 3100 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPIITTTTPTPT 3153

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 3154 TPTGTQTPTTT-PIITTTTPTPT 3175

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2528 VTPPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPT 2587

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2588 TTPITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQT-PTTPTITTTTPTPT 2646

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2647 PTPGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTPI 2706

Query: 269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLSLIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2707 TTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPT 2763

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2764 PTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIIT 2823

Query: 386 TNNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2824 TTTVTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTGT 2882

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2883 QTPTTTTPIITTTTPT-----PTPTPTGTQTPT-----TPTITTTTPTPTPTPTG--TQTP 2931

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2932 TTTPIITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTG--TGTQTPTTTPIITTTTPT 2989

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2990 PTPTPTGTQTPTTTTPIITTTTPTPTPTGTQ-TPTTPTITTTTPTPTPTGTQTPTT 3048

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 3049 TPIITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPT 3108

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 3109 PTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIIT 3168

Query: 729 AAASPPSQPAVALSTIPGAVPITPPIITIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788

P+ T P PIT TT P P+ T G+ + P V
Sbjct: 3169 TTTVTPTPTGTQT-PTTTPIT---TTTVPPTPTPT--GTQTPTTPTITTTTVPPTP 3222

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P+ P T TV+P+
Sbjct: 3223 TPTGTQTPTT-PITTTTVPPT 3244

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + TV T P TP + + P P PT P
Sbjct: 3080 VTPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPT 3139

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3140 TPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQT-PTTPTITTTTVPPT 3198

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3199 PTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTI 3258

Query: 269 IHQIQSRPPVTTSNAPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3259 TTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVP---PTPT 3315

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3316 PTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITT 3375

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTS 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3376 TTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGT 3434

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSDNRSPVQFQYEL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3435 QTPTTPTITTTTVP-----PTPTPTGTQTPT-----TPTITTTTVPPTPTPTG--TQTP 3483

Query: 503 TYTPITSSVS-TIROQPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT TV T Q T
Sbjct: 3484 TTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPT--TGTQTPTTPTITTTTVP 3541

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQQP- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 3542 PTPTPTGTQTPTTPTITTTTVPPTPTPTGTQ-TPTTPTITTTTVPPTPTPTGTQTPTT 3600

Query: 614 -PEGKTSAVVLADGATIVANPISNFFSAAPAT-TVVQTHSQSASTNAPAGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 3601 PTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPT 3660

Query: 672 RKKPATDGAKEPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 3661 PTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITT 3720

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTAAAPPSVTVGSSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 3721 TTTVPPTPTPTGTQT-PTTTPIT---TTTVPPTPTPT--GTQTPTTPTITTTTVPPTPT 3774

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P+ P T TV+P+
Sbjct: 3775 TPTGTQTPTT-PITTTTVPPT 3796

Score = 313 (47.0 bits), Expect = 4.2e-23, P = 4.2e-23
Identities = 169/695 (24%), Positives = 245/695 (35%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + TV T P TP + + P P PT P
Sbjct: 3655 VTPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPT 3714

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3715 TPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQT-PTTPTITTTTVPPT 3773

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3774 PTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTI 3833

Query: 269 IHQIQSRPPVTTSNAPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3834 TTTTVPPTPTPTGTQTPTTPTITTTTVPPTPTPTGTQTPTTPTITTTTVP---PTPT 3890

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3891 PTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIIT 3950

Query: 386 NTII-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3951 TTTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPTGT 4009

Query: 444 RIQPDYPAERSSLIPISGHRASPNPAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 4010 QTPTTTTPIITTTTIVT-----PTPTPTGTQPT-----TTPITTTTIVTPTPTPTG--TQTP 4058

Query: 503 TYTPITSSVS-TIROYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 4059 TTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPTG--TGTQPTTTTPIITTTTIVT 4116

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 4117 PTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ-TPTTTPITTTTIVTPTPTPTGTQPT- 4174

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAGSSPRPSILRKK 674
T+ + T+ P P T ++ ++N P + S+P+ S
Sbjct: 4175 ---TTPITTT--TTVTPTPTPTGTQTPPTHTSTAPIAELTTSNPPESSTPQTSRSTSS 4229

Query: 675 PATDGAKPKSEIH--VSMATPVTVSMETVSNQNDQPTIAVPP-TAQQPP--PTIPTMIA 729
P T+ S + M+ S T + T++ PP T PP PT T
Sbjct: 4230 PLTESTLLSLTPPALEMTSTAPPSTPTAPTTSGGHTLSPPPSTTSSPGTPTRGTTTG 4289

Query: 730 AASPPSQPAVALSTI---PGAVPITPP--ITTIAAAP-PPSVTVGSSLSSVLGPPVPEI 782
++S P+ V +T P P++ P I T P P SV + L+ P E+
Sbjct: 4290 SSSAPTPSTVQTTTSAWTPPTPLSTPSIIRTGLRYPSSVLICCVLNDYYAPGEEV 4349

Score = 279 (41.9 bits), Expect = 1.8e-19, P = 1.8e-19
Identities = 138/540 (25%), Positives = 194/540 (35%)

Query: 278 PVTTNSAIPPAVVATVSATRAQSPVITTTAAH-----ATDSALSRP--TLSIQHPPSAA 329
P+TT+ + P T + T +P+ TTT T + + P T + P
Sbjct: 1946 PITTITTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPT 2005

Query: 330 ISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILAT 386
Q P + TT P+ GT + T + T TP T PI T
Sbjct: 2006 TGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTT 2065

Query: 387 NTII-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSPR 444
T+ P+ T G+ + T P +T T+T P+ + T TT V P T T +
Sbjct: 2066 TTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ 2124

Query: 445 IQPDYPAERSSLIPISGHRASPNPAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAHT 503
P ++ + +P P +T + + P+ + PT P+ T
Sbjct: 2125 TPTTTTPIITTTTIVT-----PTPTPTGTQPT-----TTPITTTTIVTPTPTPTG--TQTP 2173

Query: 504 YTPITSSVS-TIROYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 561
TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2174 TTPITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTG--TGTQPTTTTPIITTTTIVT 2231

Query: 562 QPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP-- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2232 TPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ-TPTTTPITTTTIVTPTPTPTGTQPTTTT 2290

Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSILR 672
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2291 PITTITTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPT 2350

Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMIA 729
T P + TP +T + T P PT Q P T P
Sbjct: 2351 TGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTT 2410

Query: 730 AASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGSSLSSVLGPPVPEIKVKEEVE 789
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 2411 TTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTG--TGTQPTTTTPIITTTTIVTPTPT 2464

Query: 790 PMDIMRPVSAVPPPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 2465 PTGTQPTTTTPIITTTTIVTPT 2485

Score = 265 (39.8 bits), Expect = 5.8e-18, P = 5.8e-18
Identities = 179/746 (23%), Positives = 257/746 (34%)

Query: 96 VVVRYPVQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3678 VTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPT 3737

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 3738 TTPITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQT-PTTPTITTTTPTPT 3796

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + L +++ +TT T T P I
 Sbjct: 3797 PTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTI 3856

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P
 Sbjct: 3857 TTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTPTPT 3913

Query: 329 AISIQRFPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 3914 PTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITT 3973

Query: 386 TNTI-PSATTAGSVSHTOAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTS 443
 T T+ P+ T G+ + T P +T T+ P+ + T TT V P T T
 Sbjct: 3974 TTTVPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTPTPT 4032

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 4033 QTPTTPTITTTTPTPTPTPTGTQTPTPTPTITTTTPTPTPTPTPTPTPTPTPTPT 4081

Query: 503 TYTPTSSVS-TIRQPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 4082 TTTPTITTTTPTPTPTPTGTQTPTTPTITTTTPTPTPTPTPTPTPTPTPTPTPTPTPT 4139

Query: 561 IQPAPISTQGIQAPIGTPIGFI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 4140 PTPTPTGTQTPTTPTITTTTPTPTPTGTQT-TPTTPTITTTTPTPTPTPTGTQTGP 4198

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPA---ATTVVQTHSQSA-STNAPA--QGSSPRP 668
 TS +A+ T +NP P S+ P +T+ T S + ST PA S+ P
 Sbjct: 4199 T-HTSTAPIAELTT--SNP--PPESSTPQTSRSTSSPLTESTLLSTLPPAIEMTSTAPP 4253

Query: 669 SILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMI 728
 S T G S + +P + ++ PT + T T PT
 Sbjct: 4254 STPTAPTTSGGHTLSPPPSTTTSPPGTPTRGTTTGSSSAPTSTVQTTTSAWT-PTPT 4312

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 ++P L P +V I + AP V G+ + E
 Sbjct: 4313 PLSTPSIIRTGLRPYSSSVLICCVLNDYYAPGEEV-YNGTYGDTCYFVNCSLSCTLEF 4371

Query: 789 EPMIDMRPVSAPPLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQOH 841
 S P + +T +PS ++ S PT P P P +Q++
 Sbjct: 4372 YNWSCTSPSTPTPTPSKSTPTPSKP--SSTPSKPTPGTKPPECDFDPPRQEN 4422

Score = 254 (38.1 bits), Expect = 8.7e-17, P = 8.7e-17
 Identities = 167/697 (23%), Positives = 245/697 (35%)

Query: 115 SATPVAVTAPPAHLTPAVPLSFSEGLMKPPPK--PTMPSR-PIAPAPPSTLSLPPKV-PG 170
 S + T PP TP+ P + + PPP P+ P+ PI P P ST +LPP P
 Sbjct: 1587 SPPTITTTTTPPTTTTSPPTTTT---TPPTTTTSPPTTTTPTP-PTSTTTLPPTTTPS 1642

Query: 171 QVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHL 230
 T + P + P T + + TT I + P PP +
 Sbjct: 1643 PPPTTTTTPPTTTTSPPTTTTTPPTTTTTPPTTTTTPSSPI--TTTSPPTTTMTTTPS 1700

Query: 231 PRGAAAAVMSSSKVTTVLRPTSQLPNAATAQPAVQHIHQPIQS-RPPVTTNSAIPPAV 289
 P SS +TT P+S + P P + PP TT +PP
 Sbjct: 1701 P-----TTTSPSPITTTTTPSS---TTTSPPTTMTTTPSPPTTTPPTTMTLPTT 1751

Query: 290 VATVSATRAQSPVITT-TAAHATDSALSRPTLSIQH---PPSAAISIQRFPAQSRDVTTR 344
 ++ T P IT T + + + P + + + S + +P ++
 Sbjct: 1752 TSSPLTTPLPSITPPTFSFSTTTTPTCPLCNWTGWLDGKPNFHKPGGDTELGID 1811

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATN-----TIPSATTAGS 397
 + P A + + ++ I G V ++ N I P A
 Sbjct: 1812 VCGPGWAANISCRATMYP--DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNY 1869

Query: 398 VSHTOAPTSTI--VTMTVPSHSSHATAVTTSNIPVAKVVPQQITHTSPPRIQPDYPAERSS 455
 + Q TMT + + + T TT+ I V T T + P ++
 Sbjct: 1870 EINVQCCCVTQPTTMTTTT-TENPTPTTTPITTTTPTPTPTGTQTPTTPTITTTT 1928

Query: 456 LIPISGHRASPNPVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAHTYTPITSSVS-T 513
 + +P P +T + + P+ + PT P+ T TPIT++ + T
 Sbjct: 1929 TVT-----PTPTPTGTQTPT---TTPITTTTPTPTPTPTG--TQTPTTPTITTTTPT 1977

Query: 514 IRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQAPAPISTQGIQ 572

Score = 243 (36.5 bits), Expect = 1.3e-15, P = 1.3e-15
Identities = 110/406 (27%), Positives = 154/406 (37%)

Score = 189 (28.4 bits), Expect = 8.0e-09, P = 8.0e-09
Identities = 92/374 (24%), Positives = 133/374 (35%)

773

Query: 786 EEVEPMDIMRPVSAVPLATNTVSPSL 812
 M + P + PL T + PS+
 Sbjct: 1739 PPTTMTTLPPPTTSSPLTTTLPPLSI 1765

Score = 185 (27.8 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 71/270 (26%), Positives = 99/270 (36%)

Query: 563 PAPISTQGIQAPIGTPGIQAPPLGTQGIHSATP---INTQGLQAPMGTQQPQ---PEG 616
 P+P +T P P TP P T + + TP I+T P P T P P
 Sbjct: 1422 PSPPTTTTTPPTTTPS-PPITTTTTLPTTTPSPPISTT-TTPPTTTPSPPTTTPSP 1479

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSPPRPSILRKKPA 676
 T+ T P + P +P TT + T S +T P SP + P
 Sbjct: 1480 PTTTSPPTTTTTPPTTTP---SPPMITPI-TPPASTTTLPTTTPSPPTTTTTPPP 1535

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQ 736
 T P + TP+T T + P+ P T PPPT + PS
 Sbjct: 1536 TTTTSPPT-----TTPITPPTSTTLPTTTPS-PPPTTTTTPPTTTPSPPTTTPSP 1588

Query: 737 PAVALSTIPGAVPITPPITTIAAAPPVSVVGGSLSSVLGPPVPEIKVKEVEPMDIMRP 796
 P + +T P +PP TT PPP+ T ++ + PP + P
 Sbjct: 1589 PTITTTTTPPTTTPSPPTT-TTPPTTTPSPPTTTPITPPTSTTLPTTTPSP--PP 1645

Query: 797 VSAVPLATNTVSPSLALLANLMSPTSDLPFGASP 832
 + P T T SP + T+ PP +P
 Sbjct: 1646 TTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTP 1681

Score = 183 (27.5 bits), Expect = 3.4e-09, P = 3.4e-09
 Identities = 91/390 (23%), Positives = 139/390 (35%)

Query: 326 PSAAISIQRPQSRDVTTR-ITLPSHPALGTPKQQLHTMAQKTIISTGTPVAAATVAPIL 384
 PS + P + T T PS P T T I +T TP+ T +P +
 Sbjct: 1399 PSPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPIITTTTTLPTTTPSPPI 1458

Query: 385 ATNTIPSATTAGSVSHTQAPTSTIVMTVPSHSHATAVTTSNIP--VAKVVPQQITHTS 442
 +T T P TT S T P+ T +P+ ++ TT+ P + P T T
 Sbjct: 1459 STTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPITPPASTTTL 1517

Query: 443 PRIQPDYPAERSSLPISGHRASP---NPVAMETRSNRP--SVPVQFYFLPTYPPSAY 497
 P P ++ P SP P+ T + P + P T PP+
 Sbjct: 1518 PTTTTPSPPTTTTTPPTTTPSPPTTTPITPPTSTTLPTTTPSPPTTTPPTTTPPTT 1577

Query: 498 PLAHAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTVGVASTVHLNPMQL-MTVDASHAR 556
 P T TP ++T P + +P T T +T P +T S
 Sbjct: 1578 PSPPTTTPSPPTITTTTTPPTTTPSPPTTTPPTTTPPTTTPITPPTSTTT 1634

Query: 557 HIQGIQAPISTQGIQAPIGTPGIQAPPLGTQGIHSATPINTQGLQAPMGTQQPQPEG 616
 P+P T P P TP P P T T T P P
 Sbjct: 1635 LPPTTTPSPPTTTTTPPTTTPS--P-PTTTPSPPIITTTTTPPTTTPSSPITTTTPSP 1691

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSPPRPSILRKKPA 676
 T+ + T ++PI+ + P+TT + +T +P SP + + P
 Sbjct: 1692 PTTTMTTPSPTTTPSSPITT--TTTSSSTTPSPPTTMTTPSPTTTPSPTTMTTLPP 1749

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPP 715
 T + P + + P +++ TS + PT P
 Sbjct: 1750 TTTSSPLT----TTLPPSITPPTTFSPFSTTTPTTPCVP 1784

Score = 176 (26.4 bits), Expect = 1.8e-07, P = 1.8e-07
 Identities = 101/402 (25%), Positives = 142/402 (35%)

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIISTGTPVAAATVAPILATNTIPSATTAGSVSHTQAP 404
 IT PS P TP T +T +P T P T P TT + T P
 Sbjct: 1396 ITTSPPTT-TPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPIITTTTTLPTTTP 1454

Query: 405 TSTIVMTVPSHSHATAVTTNIPVAKVVPQQITHSPRIQPDYPAERSSLPISGHR 463
 + I T T P ++ + TT+ + P P T T+P P PI+
 Sbjct: 1455 SPPITTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTP--PPTTTPSPMTTTPITPP- 1511

Query: 464 ASPNPVAMETRSNRPVQFYFLPTYPPSAYPLAHTYTPITSSVSTIRQYPVSAQA 523
 AS + T PS P T PP+ P + T TPIT ST P + +
 Sbjct: 1512 ASTTTLPTTT----PSPPTTTT--TTPPTTTP-SPPTTTPITPPTSTTLPTTTPS 1563

Query: 524 PNSAITAQ----TGVGVASTVHLNPMQLMTVDASHARHIQGIQAPISTQGIQAPIGTP 579
 P T T T +T +P + T P+P +T P P TP
 Sbjct: 1564 PPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTP--TPSPPTTTTTPPTTTP 1618

Query: 580 G-----IQAPPLGTQGIHSAT---PINTQGLQAPMGTQQPQPEGKTSAVVLADGATIV 630
 I P P T + T P T P P T P S +
 Sbjct: 1619 SPPTTTPITP-PTSTTTLPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTP 1677

Query: 631 ANPISNPFSAAPAA-TTVVQTHSQSASTNAP-AQGSSPRPSILRKKPATDGAKPKSEIHV 688
 S+P + P+ TT + T S + + +P +P + P T P
 Sbjct: 1678 TTTSSPITTTSPPTTTMTTPSPTTTPSSPITTTTTPSSSTTTSPSPPTTMTTPSP---T 1734

Query: 689 SMATPVTVMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPG 746
 + +P T +M T+ P P PPT + + P+ P V L G
 Sbjct: 1735 TTPSPPTTMTTLPPTTSSPLTTTLPSPITPPTSPF--STTTPTPCVPLCNWTG 1790

Score = 168 (25.2 bits), Expect = 9.3e-08, P = 9.3e-08
 Identities = 89/387 (22%), Positives = 133/387 (34%)

Query: 448 DYPAERSSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFLPTYPPSAYPLAAHTYTP 507
 DY + P+ +P+ T + + P P PT PS P T P
 Sbjct: 1381 DYKIRVNCCWPMKDCITTPSP---PTTTPSP--PTTTTLPPTTTPSP-PTTTTTTPPP 1434

Query: 508 TSSVS---TIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQPA 564
 T++ S T P+ P+ I+ T +T P T + P+
 Sbjct: 1435 TTTSPPTITTTTLPPTTTPSPPTTTTTPPTTT---PSPTTTPSPPTT-----TPS 1485

Query: 565 PISTQGIQPAPIGTPIGI-QPAPLGTQGIHSATPINTQGLQPAPMGTOQFQ---PEGKTS 620
 P +T P P TP P+ + P T P T P P T+
 Sbjct: 1486 PPTTTTTPPTTTPSPPTTTPITPPASTTTLPTTTPSPPTTTTTPPTTTPSPPTT 1545

Query: 621 VVLADGATIVANPISNPFSAAPAAATTVVQTHSQSA-STNAPAQGS---SPRPSILRKKP 675
 + +T P + P TT T + S +T P+ + +P P+ P
 Sbjct: 1546 PITPPTSTTTLPTTTPSPPTTTTTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSP 1605

Query: 676 ATDGAKPKSEIHVS--MATPVTVMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASP 733
 T P S TP+T T + P+ P T PPPT +
 Sbjct: 1606 TTTTTTPPTTTPSPPTTTPITPPTSTTTLPTTTPS-PPPTTTTTPPTTTPSPPTT 1664

Query: 734 PSQPAVALSTIPGAVPITPPITIAAAPPSVTVGSLSSVLGP---PVPEIKVKEEVE 789
 PS P +T P + PITT + P ++T ++ P P
 Sbjct: 1665 PSPPTTTTTPPTTTPSPPTTTPSPPTTMTTPSPPTTTPSSPITTTTTPSSTTTPSP 1724

Query: 790 PMDIRPVSAPVPLATNTVSPSLALLANLMSMPTSDLPFGASP 832
 P + P P T +L + + T+ LPP +P
 Sbjct: 1725 PTTMTTPSPPTTTPSPPTTMTTLPPTTSSPLTTTLPSPITP 1767

Score = 154 (23.1 bits), Expect = 2.7e-06, P = 2.7e-06
 Identities = 70/277 (25%), Positives = 92/277 (33%)

Query: 565 PISTQGIQPAPIGTPIGIQAPLGTQGIHSATPINTQGLQPAPMGTOQFQPEGKTS 624
 PIST P P TP P P T + TP P T P P T +
 Sbjct: 1457 PISTT-TTPPTTTPS--P-PTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTP--ITP 1510

Query: 625 DGATIVANPISNPFSAAPAAATTVVQTHSQSASTNAP---AQGSSPRPSILRKKPATDGA 680
 +T P + P TT T + S T P ++ P+ P T
 Sbjct: 1511 PASTTTLPTTTPSPPTTTPPTTTPSPPTTTPITPPTSTTTLPTTTPSPPTT 1570

Query: 681 KPKSEIHVSMATPVTVMETVSNQNDQPTIAVPPTAQ--PPPTIPTMIAAASPPSQPA 738
 P S T T S T++ T PPT PPPT T + P P
 Sbjct: 1571 TTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSPPTTTPPTTTPSPPTTTPSPPTT 1629

Query: 739 VALSTIPGAVPITPPITIAAAPPSVTVGSLSSVLGPVPEIKVKEEVEPMDIRPVS 798
 + +T+P +PP TT PPP+ T ++ PP+ +
 Sbjct: 1630 TSTTTLPPTTTPSPPTT-TTTPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTT 1688

Query: 799 AVPPLATNTV----SPSLALLANL--SMPTSDLPFGASPRKKP 836
 PP T T +PS + S T PP P
 Sbjct: 1689 PSPPTTMTTPSPPTTTPSPPTTTPSSTTTPSPPTTMTTPSP 1733

Score = 148 (22.2 bits), Expect = 1.1e-05, P = 1.1e-05
 Identities = 62/254 (24%), Positives = 89/254 (35%)

Query: 583 PAPLGTQGIHSATPINTQGLQPAPMGTOQFQPEGKTSAV----VLADGATIVANPISNP 637
 P+P T S P T L P T P P T+ + T P+
 Sbjct: 1399 PSPPTTTP--SPPTTTTTLPP---TTTPSPPTTTPPTTTPSPPTTTPPTTTP 1452

Query: 638 FSAAPAAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKPKSEIHVS--MATPVT 695
 + P +TT T + + + P SP P+ P T P S M TP+T
 Sbjct: 1453 TSPPTTITTT--TPPTTTPSPPTTTPSP-PTTTPSPPTTTPPTTTPSPPTTTP 1509

Query: 696 VSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPIT 755
 T + P+ T PP T P+ + P P + +T+P +PP T
 Sbjct: 1510 PPASTTLPPTTTPSPPTTTPPTTTPPTTTPS--PPTTTPITPPTSTTTLPTTTPSPPT 1567

Query: 756 TIAAAPPSVTVGSLSSVLGPVPEIKVKEEVEPMDIRPVSAPVPLATNTVSPSLALL 815
 T PPP+ T ++ PP + PP T P+ +
 Sbjct: 1568 T-TTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTP 1626

Query: 816 ANNLSMPTSDLPPGASPRKKP 836
 S T+ LPP +P P
 Sbjct: 1627 TPPTS--TTTLPPTTTSPSP 1645

Score = 131 (19.7 bits), Expect = 1.2e-03, P = 1.2e-03
 Identities = 112/492 (22%), Positives = 174/492 (35%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAUTAPPAHL-TPAVPLSFSEGLMKPPPKTTPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3977 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 4036

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGOQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 4037 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 4095

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAA-VMSKKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 4096 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 4155

Query: 269 IHQPIQSRPVTTSNAIPPA--VVATVSATRAQSPVITTA--AHATDSALSRLTSLIQH 324
 + P T P + T + T +P T T H + + + T S
 Sbjct: 4156 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTGPPTHTSTAPIAELTTSNPP 4215

Query: 325 PPSAAISIQRPQS--RDVTTRI-TLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVA 381
 P S+ R S + TT + TLP PA+ + T T + T T++
 Sbjct: 4216 PESSTPQTSRSTSSPLTESTTLLSTLP--PAI----EMTSTAPPSTPTAPTTTSGGHTLS 4269

Query: 382 PILANTIPSAT-TAGSVS-HTQAPTSTIVMTVPSSHSHATAVTTSNIPVAKVVPQKIT 439
 P +T T P T T G+ + +APT + V T S A T + P++ P I
 Sbjct: 4270 PPSPTTSPPGTPTGTGTGSSSAPTSTVQTTTS-----AWTPTPTPLS--TPSIIR 4321

Query: 440 HTSPRIQPDYPAERSSSLIPISGHRASPNP-VAMETRSDN----RPSVPVQFQYFLPTYP- 493
 T ++P YP+ ++ +P V T D S+ +++ + P
 Sbjct: 4322 TTG--LRP-YPSSVLICCVLNDYYAPGEEVYNGYGDTCYFVNCSLSCTLEFYNWSCPS 4378

Query: 494 -PSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVAHVHLPMLMTVDA 552
 PS P + + TP S S+ P P T L + T
 Sbjct: 4379 TPSPTPTPSKS-TPTSPKPSSTPSKPTPGTKPECPDFDPPRQENETWWLDCDFMATCKY 4437

Query: 553 SHARHIQGIQ---PAPISTQGIQAPIGTP 579
 ++ I ++ P P + G+QP + P
 Sbjct: 4438 NNTVEIVKVECEPPMPPTCSNGLQPVVRVEDP 4468

Score = 117 (17.6 bits), Expect = 1.8e-02, P = 1.8e-02
 Identities = 41/156 (26%), Positives = 55/156 (35%)

Query: 710 TIAPVPTAQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITIAAAPPSPVTVGG 769
 T + P T PPPT T + + PS P +T P +PPIIT P P+ T
 Sbjct: 1398 TPSPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPIIT-TTTLPTTTPSP 1456

Query: 770 SLSSVLGPPVPEIKVKEEVEPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPPG 829
 +S+ PP P P + P T T SP T+ PP
 Sbjct: 1457 PISTTTTP-----PTTTPSPPTTTPSPPTTTPSPPTTTTTP-PPTTTPSPPM 1504

Query: 830 ASPRKKPRKQHVISTEEDMMETNSTDDEKSTAKS 865
 +P P + T T+T +T S
 Sbjct: 1505 TTPITPPASTTTLPTTTPSPPTTTTTPPTTTPS 1540

Score = 61 (9.2 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 23/93 (24%), Positives = 41/93 (44%)

Query: 397 SVSHTQAPTSTIVMTVPSSHSHATAVTTSNIPVAKVV---PQIHTSPRIQPDYPAE 452
 S++ + +T T+T+P+ + T TT+ P + V P+ S I D+P+
 Sbjct: 1257 SITTRPSTLTFTTTITLPTTPTSTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSS 1316

Query: 453 RSS---LIPISGHRASPNPVAMETRSDNRPSVPVQ 484
 S P G +P + E RS P + +
 Sbjct: 1317 GSDDGDREPFQVCGAPEDI--ECRSVKDPHLSLE 1349

Score = 50 (7.5 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 16/41 (39%), Positives = 19/41 (46%)

Query: 334 RPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTP 374
 RP+ TT ITLP+ P T T T+ ST TP
 Sbjct: 1261 RPSTLTFTTT-ITLPTTPTSTTTTTTTTPTSSTVLST-TP 1299

Score = 46 (6.9 bits), Expect = 5.4e-08, P = 5.4e-08
 Identities = 24/106 (22%), Positives = 37/106 (34%)

Query: 324 HPPSAAISIQRPQASRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPI 383
 +PP A++ + +S T + P G Q A G I

Sbjct: 1196 YPPGASVPTEETCKSCVCTNSSQVVCREEGKILNQTQDGAFCYWEICGPNGTVEKHFN 1255

Query: 384 LATNTIPSA-TTAGSVSHTQAPTSTIVTMTVPSSHSHATAVTTSNI 428
+ T P S T T + + + P T S T T + + S T T +

Sbjct: 1256 CSITTRPSTLTFTTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKL 1301

Score = 44 (6.6 bits), Expect = 8.7e-08, P = 8.7e-08
Identities = 14/34 (41%), Positives = 17/34 (50%)

Query: 478 RPSVPVQFQYF-LPTYPPSAYPLAAHTYTPITSSV 511
RPS F LPT P S + T TP +S+V

Sbjct: 1261 RPSTLTFTTTITLPTTPTS-FTTTTTTTTPTSSTV 1294

Pedant information for DKFZphtes3_2a11, frame 2

Report for DKFZphtes3_2a11.2

[LENGTH] 1048
[MW] 110324.04
[pI] 9.83
[HOMOL] PIR:I47141 gastric mucin (clone PGM-2A) - pig (fragment) 8e-15
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR420w] 4e-09
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR420w] 4e-09
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR151c] 4e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGR014w] 1e-05
[FUNCAT] 11.01 stress response [S. cerevisiae, YHL028w] 1e-04
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YHL028w] 1e-04
[EC] 3.2.1.3 Glucan 1,4-alpha-glucosidase 3e-08
[PIRKW] glycosidase 3e-08
[PIRKW] transmembrane protein 3e-08
[PIRKW] polysaccharide degradation 3e-08
[PIRKW] glycoprotein 9e-08
[PIRKW] calcium binding 9e-08
[PIRKW] hydrolase 3e-08
[PIRKW] cytoskeleton 7e-08
[SUPFAM] equine herpesvirus glycoprotein X 2e-07
[SUPFAM] yeast glucan 1,4-alpha-glucosidase homolog 3e-08
[SUPFAM] polymorphic epithelial mucin 7e-08
[SUPFAM] glucan 1,4-alpha-glucosidase homology 3e-08
[SUPFAM] equine herpesvirus 1 glycoprotein homology 2e-07
[PROSITE] MYRISTYL 9
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 10
[PROSITE] PKC_PHOSPHO_SITE 12
[PROSITE] ASN_GLYCOSYLATION 3
[KW] Irregular
[KW] LOW_COMPLEXITY 20.04 %

SEQ MGPPRHPQAGEIEAGGAGGRRRLQVEMSSQQFPRLGAPSTGLSQAPSQIANSGSAGLINP
SEGxxxxxxxxxxxxx.....
PRD ccc

SEQ AATVNDSEGRDSEVSAREHMSSSSSLQSRREEKQEPVVVRPYPQVQMLSTHHAVASATPVA
SEGxxxxx.....xxxxxxxxxxxxx
PRD ccc

SEQ VTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPQGQVTVTMESSI
SEG xxxxxxxxxxxxxxxx.....xxxxxxxxxxxxx
PRD ccc

SEQ PQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNA GPPLHIGASHLPRGAAAAAVM
SEGxxxxx..
PRD ccc

SEQ SSSKVTTVLRPTSQLPNAATAQPAVQHIIHQPIQSRPPVTTNSAIPPAVVATVSATRAQS
SEG
PRD ccc

SEQ PVITTTAAHATDSALSRPTLSIQHPPSAAISIQRPASRDVTTTRITLPSHPALGTPKQQL
SEG
PRD ccc

```

SEQ HTMAQKTI FSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAPTSTIVTMTVPSSHSHA
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TAVTTSNIPVAKVVPQQIHTSPRIQPDYPAERSSSLIPISGHRASPNPVAMETRSDNRPS
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ VPVQFYFLPTYPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTVGVVASTV
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPIGTGPIQAPPLGTQGIHSATPINTQ
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GLQPARMGTTQQPQEGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AQGSSPRPSILRKKPATDGAKPKSEIHVSMATPVTVMETVSNQNNDOPTIAVPPTAQQP
SEG .....XXXXXXXXXX
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PPTIPTMIAAASPSPQPAVALSTIPGAVPITPPITIAAAPPSVTVGGLSSVLGPPVP
SEG .....XXXXXXXXXX
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EIKVKEEVEPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPGASPRKKPRKQO
SEG .....XXXXXXXXXX
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HVISTEEGDMETNSTDDEKSTAKSLLVKAERKSPPEYIDEEGVRYVPVVRPPTITLL
SEG .....XXXXXXXXXX
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RHYRNPWKAAYHHFQRYSDVRVKEEKAMLEIATNQKGVSCRAQGWKVLCAAQLQLTN
SEG .....
PRD eccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LEHDVYERLTNLQEGIIPKKAATDDDLHRINELIQNMQRCKLVMDQISEARDSMLKVL
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DHKDRVLKLLNKGTVKKVSKLRKEKV
SEG .....XXXXXXXXXX
PRD hhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_2all.2

PS00001	818->822	ASN_GLYCOSYLATION	PDOC00001
PS00001	854->858	ASN_GLYCOSYLATION	PDOC00001
PS00001	1033->1037	ASN_GLYCOSYLATION	PDOC00001
PS00004	872->876	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1037->1041	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005
PS00005	242->245	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	665->668	PKC_PHOSPHO_SITE	PDOC00005
PS00005	831->834	PKC_PHOSPHO_SITE	PDOC00005
PS00005	862->865	PKC_PHOSPHO_SITE	PDOC00005
PS00005	940->943	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1035->1038	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	88->92	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	473->477	CK2_PHOSPHO_SITE	PDOC00006
PS00006	844->848	CK2_PHOSPHO_SITE	PDOC00006
PS00006	855->859	CK2_PHOSPHO_SITE	PDOC00006
PS00006	959->963	CK2_PHOSPHO_SITE	PDOC00006
PS00006	984->988	CK2_PHOSPHO_SITE	PDOC00006
PS00008	15->21	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	16->22	MYRISTYL	PDOC00008
PS00008	36->42	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	535->541	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	768->774	MYRISTYL	PDOC00008
PS00009	19->23	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_2a11.2)

DKFZphtes3_2a17

group: metabolism

DKFZphtes3_2a17 encodes a novel 574 amino acid protein without similarity to known proteins.

The novel protein contains a thiol protease cys pattern. Eukaryotic thiol proteases (EC 3.4.22.-) are a family of proteolytic enzymes containing an active site cysteine. Cathepsins belong to this protease family.

The new protein can find application in modulation of proteolytic processes and as a new enzyme for proteomic analysis and biotechnologic production processes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 2312 bp

Poly A stretch at pos. 2300, polyadenylation signal at pos. 2273

```
1 GTTTTCACCT GATCATTAGA AACTAATGAA ACACCTTTTA AGTCTTATGA
51 ATTCAGGTTA CACTGTTTTC CAGATGCCCTT GGCAGCTGGT ACAGGGCCTC
101 TGA AAAAATGG AACCAAATTC TCTGAGGACT AAAGTCCCAG CTTTCTTATC
151 TGATTTGGGG AAGCCACAT TGAGGGGAAT CAGAAAGTGT CCCCGATGTG
201 GCACATACAA TGGAAACCCGG GGACTGAGCT GTAAGAACAA GACATGTGGA
251 ACCATATTCC GCTACGGTGC ACGCAAGCAG CCTAGTGTG AAGCTGTCAA
301 AATCATTACA GGCTCTGATC TTCAGGTCTA CTCAGTGGCG CAAAGAGACC
351 GGGGCCCTGA TTACCGATGC TTTGTGGAGC TCGGGGTTTC AGAGACAACA
401 ATCCAGACAG TGGATGGGAC GATCATCACT CAGCTGAGCT CTGGACGGTG
451 TTATGTCCCC TCATGCCTGA AAGCTGCCAC TCAAGGCGTT GTGGAAAACC
501 AGTGCCAGCA CATCAAGCTG GCGGTGAACT GCCAGGCAGA GGCCACCCTT
551 CTGACCTCGA AGAGCTCGGT CCTGAATGCA ATGCAGGCCT CCCCGAAAC
601 CAAACAGACC ATCTGGCAGT TGGCCACGGA ACCCAGAGGT CCTCTGGTGC
651 AGAGAATTAC TAAAAACATC TTGGTGGTGA AATGCAAGGC AAGCCAGAA
701 CACAGTTTGG GGTATTTGCA TACATCTTTT GTGCAGAAAG TCAGTGGCAA
751 AAGCTTGCCT GAGCGCCGCT TCTTCTGCTC CTGTCAAGCT CTGAAATCGC
801 ACAAGTCAAA TGCTTCCAAG GATGAGACAG CCCAGAGATG CATTCAATTC
851 TTTGCTTGCA TCTGTGCTT TGCCAGTGAT GAGACACTGG CTCAGGAATT
901 CTCAGACTTC CTAATTTTG ATTCCAGCGG TCTTAAAGAG ATTATGTAC
951 CCCAGTTAGG TTGCCATTCA GAATCAACAG TATCTGCTTG TGAGTCTACT
1001 GCCTCTAAGT CAAAGAAGAG GAGAAAGGAT GAAGTATCTG GTGCACAGAT
1051 GAACAGTTCA CTACTGCCTC AAGATGCAGT GAGCAGTAAT CTAAGGAAAA
1101 GTGGCCTGAA AAGCCTGTG GTTGCTTCC TTTAAAAAG GCAGGCCTGT
1151 GGTGAGCTGT TAGATGAGGC ACAAGTGACT TTATCCTTCC AAGACTGGCT
1201 GGCCAGTGTC ACAGAACGCA TCCATCAAAC CATGCACTAT CAGTTTGATG
1251 GCAAACACAGA ACCATTGGTG TTCCACATTC CTCAGTCATT TTTGATGCC
1301 CTGCAACAAA GAATATCTAT AGGAAGTGCA AAAAAACGGC TCCCAACTC
1351 CACCACAGCT TTTGTTGCGA AAGATGCCCT GCCACTGGGA ACCTTTTCCA
1401 AGTATACTTG GCATATCACT AATATCCTGC AAGTTAAACA AATCTTAGAT
1451 ACCCCAGAGA TGCCCTTGGA AATCACCCGT AGCTTTATCC AGAACCAGAG
1501 TGGGACTTAT GAGCTATTTA AATGCCCTAA AGTGGAAGTA GAAAGCATAG
1551 CAGAAACCTA CGGTCTGATA GAAAAACAAC CAGTGCTGCG ACCCTTGGAA
1601 CTA AAAA ACTT TTCTCAAAGT TGGCAACACT TCCCAAGATC AAAAGGAGCC
1651 AACACCTTTC ATCATCGAGT GGTATCCAGA TATCCTTCCC CAATCTAAGA
1701 TTGGCGAGCT GCGGATCAAG TTTGAGTATG GCCACCACCG GAATGGGCAT
1751 GTGGCGGAGT ACCAAGACCA GCGGCCCCCC TTGGACCAGC CCTTGGAACT
1801 GGCCCTCTCT ACCACTATTA CTTTCCCTTA AAGCAAAACA AGATAATAAT
1851 CTTTTGCTGC TTAATTTGCA CATCCCCACC CCTTGACAAC TTTAATGCT
1901 AGTTAGGCAC TTAGATGGCC CTGTTCCTTG GTAACTGCT CTTAGCTAAG
1951 ATGCAAAATC TCAGTGCTTT CAAGTGGATT CTGTTGAAGA AATCTCTTG
2001 TAAATAGCCT TTTGATGCT GCTGTGTACA GTCTTCATTA TGCATTGGGC
2051 AGTATTTCTG GCTAGAGTTT TAAAAGGAAC AGAAAGAAAA CCAGCTTATT
2101 TTCCTTCTTA CGGACTCATC TTAGCGTTT ATTCAACCT TTTGCTAATT
2151 CTCTGAGAAA TCTGCAGCAC TCAGCCATAC ACCAACAGTG TTGGAAAAGT
2201 AACACCTTGG TTAGGGCAGA ATGTTAAAGA CCATCTTGGC AGAGTTCCAG
2251 CCACGCTCTT TATTCTGTTT TCAAATAAAG CAGTGTCACT AGTTTTCTCT
2301 AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 1828 bp; peptide length: 574
 Category: putative protein

```

1  MEPNSLR TKV  PAF LSDLGKA  TLRGIRK CPR  CGTYNGTRGL  SCKNKT CGTI
51 FRYGARKQPS  VEA VKIITGS  DLQVYSVRQR  DRGPDYRCFV  ELGVSETTIQ
101 TVDGTIITQL  SSGRCYVPSC  LKAATQGVVE  NQCQHIK LAV  NCQAEATPLT
151 LKSSVLNAMQ  ASPETKQTIW  QLATEPTGPL  VQRITKNILV  VKCKASQKHS
201 LGYLHTSFVQ  KVSQKSLPER  RFFCSCQTLK  SHKSNASKDE  TAQRCHFFA
251 CICAFAFDET  LAQEFSDFLN  FDSSGLKEII  VPQLGCHSES  TVSACESTAS
301 KSKKRRKDEV  SGAQMNSSL  PQDAVSSNLR  KSGLKKPVVA  SSLKRQACGQ
351 LLDEAQVTL  FQDWLASVTE  RIHQTMHYQF  DGKPEPLVFH  IQSFFDALQ
401 QRISIGSAKK  RLPNSTAFV  RKDALPLGTF  SKYTWHITNI  LQVKQILDTP
451 EMPLAITRSF  IQNRDGTIEL  FKCPKVEVES  IAETYGRIEK  QPVLRLPELK
501 TFLKVGNTSP  DQKEPTFFII  EWIPDILPQS  KIGELRIKFE  YGHRHRGHVA
551 EYQDQRPPLD  QPLELAPLTT  ITFP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2a17, frame 2

Report for DKFZphtes3_2a17.2

```

[LENGTH]      574
[MW]           64076.89
[pI]           9.15
[PROSITE]      MYRISTYL           5
[PROSITE]      CK2_PHOSPHO_SITE      9
[PROSITE]      PKC_PHOSPHO_SITE     14
[PROSITE]      ASN_GLYCOSYLATION     5
[PROSITE]      THIOL_PROTEASE_CYS    1
[KW]           Alpha_Beta

SEQ  MEPNSLR TKVPAFLSDLGKATLRGIRK CPRCGTYNGTRGLSCKNKT CGTIFRYGARKQPS
PRD  cccccccccchhhhhccccchhhhhcccccccccccccccccccccccccccccccccccccc

SEQ  VEA VKIITGS DLQVYSVRQRDRGPDYRCFVELGVSETTIQTVDGTIITQLSSGRCYVPSC
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ  LKAATQGVVENQCQHIK LAVNCQAEATPLTLKSSVLNAMQASPETKQTIWQLATEPTGPL
PRD  hhhhhhhhhccccchhhhhheehhhhhhhccccchhhhhhhhhccccchhhhhhhhhccccch

SEQ  VQRITKNILVVKCKASQKHS LGYLHTSFVQKVSQKSLPERFFCSCQTLKSHKSNASKDE
PRD  hhhhhhhheeeeecccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TAQRCHFFACICAFAFDET LAQEFSDFLNFDSSGLKEIIVPQLGCHSESTVSACESTAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KSKKRRKDEVSGAQMNSSLLPQDAVSSNLRKSGLKKPVVASSLKRQACGQLLDEAQVTL
PRD  ccchhhhhccccccccccccccccccccchhhhhhhccccceehhhhhhhhhchhhhhhhhh

SEQ  FQDWLASVTERI HQTMHYQFDGKPEPLVFHI PQSFFDALQQRISIGSAKKRLPNSTAFV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  RKDALPLGTFSKYTWHITNILQVKQILDTPEMPLAITRSFIQNRDGTIELFKCPKVEVES
PRD  eccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ  IAETYGRIEKQPVLRLPELKTFLKVGNTSPDQKEPTFFII EWIPDILPQSKIGELRIKFE
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

SEQ YGHRNGHVAEYQDQRPPLDQPLELAPLTTITFP
 PRD ecccccccccccccccccccccccccccccccccccc

Prosites for DKFZphtes3_2a17.2

PS00001	35->39	ASN_GLYCOSYLATION	PDOC00001
PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	235->239	ASN_GLYCOSYLATION	PDOC00001
PS00001	316->320	ASN_GLYCOSYLATION	PDOC00001
PS00001	414->418	ASN_GLYCOSYLATION	PDOC00001
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	231->234	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	369->372	PKC_PHOSPHO_SITE	PDOC00005
PS00005	407->410	PKC_PHOSPHO_SITE	PDOC00005
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	360->364	CK2_PHOSPHO_SITE	PDOC00006
PS00006	367->371	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	508->512	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	312->318	MYRISTYL	PDOC00008
PS00139	109->121	THIOL_PROTEASE_CYS	PDOC00126

(No Pfam data available for DKFZphtes3_2a17.2)

DKF2phtes3_2d15

group: testes derived

DKF2phtes3_2d15 encodes a novel 274 amino acid protein with similarity to C.elegans Cosmid F25H2.1.

The novel protein contains a Pfam predicted C2-domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans F25H2.1

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3615 bp

Poly A stretch at pos. 3603, polyadenylation signal at pos. 3578

```
1 GCGCGGGCCT CGAGGTGACA ACTGTCTCCG TCGCAGGCTC CGGCGGGGGC
51 GCAGGAGGTC GCCCGGCGCG TCACTGTCCG GTCCGCGAGC CACGGGGGGC
101 GCCGAGCAC CATGGCGACC ACCGTGACGA CTCAGCGCGG GCCGGTGATC
151 ATCGGTGAGC TCCCGCAGGA CTTCCTCCGC ATCAGCGCCA CACAGCAGCA
201 GCGCGAGGTC CAGCTGGACG CCCAGGCGGC CCAGCAGCTG CAGTACGGAG
251 GCGCAGTGGG CACCGTGGGC CGACTGAACA TCACGGTGGT ACAGGCAAAAG
301 TTGGCCAAGA ATTACGGCAT GACCCGCGAT GACCCCTACT GCCGACTGCG
351 CCTGGGCTAC GCGGTGTACG AGACGCCAC GGCACACAAT GGCGCCAAGA
401 ATCCCGCTG GAATAAGGTC ATCCACTGCA CGGTGCCCCC AGGCGTGGAC
451 TCTTTCTATC TCGAGATCTT CGATGAGAGA GCCTTCTCCA TGGACGACCG
501 CATTGCCTGG ACCCATCA CAATCCCGGA GTCCCTGAGG CAGGCAAGG
551 TGGAGGACAA GTGGTACAGC CTGAGCGGGA GGCAGGGGGA CGACAAGGAG
601 GGCATGATCA ACCTCGTCAT GTCTACGCG CTGCTTCCAG CTGCCATGGT
651 GATGCCACCC CAGCCCGTGG TCCTGATGCC AACAGTGTAC CAGCAGGGCG
701 TTGGCTATGT GCCCATCACA GGGATGCCCG CTGTCTGTAG CCCCGGCATG
751 GTGCGCGTGG CCCTGCCCCC GGC CGCGGTG AACGCCAGC CCCGCTGTAG
801 CGAGGAGGAC CTGAAAGCCA TCCAGGACAT GTTCCCAAC ATGACCAGG
851 AGGTGATCCG CTCGCTGCTG GAAGCCAGC GAGGGAACAA GGATGCCGCC
901 ATCAACTCCC TGCTGCAGAT GGGGAGGAG CCATAGAGCC TCTGCTCGA
951 TGCCGTTTTG CCCCCTCTT TGGACACGC CGACCCGCGC CTCCCAAGG
1001 AATGCTCTCC CAACAAGATT CCCGTGAAAG AGCACCCGTG TCGCCCCCTC
1051 CCGTGGACTT CTGTGCCGCC CCGTCCACAC CTGTTCTTGG GTGCATGTGG
1101 GTTTTCGGTT CCTGGCGGTC CAGGACGGGG CGGGGGCTCC CCTCCATCT
1151 CGTCTGGGA GGTCTCAGCG CGCTCTCCTG TCCTGGGAC GTGCGTCTCT
1201 CCTTCTCATG CCGTTCTGGA AAATGCTCTT GCTGTAGAGA GCAGCTGCTT
1251 CTGCCAGGGT GTTGGAGGTG GTGGAGCGCC TTCCGATTCC ATTCTATGCA
1301 TTTTGTGATG TGATGTAATT GGAATAGAGC TGTGATTTA AGGCACACAC
1351 AATCCCTCAC ACTGTGGGTT TTTTITAGAA CTTCACAGAC GAAACTCAC
1401 GCCCTTGCCC TAACGCGCTT TGCTGTGAGC CTGGCCCTTG CCCAGGGCTT
1451 GGGCTGTGGT AGCTGAGCAG CTCTCTGTGG ATGGTGTGGG GCCGCGCTCT
1501 GGCCTGGCTC ACCTGGCCAC TGTCCAGCCA GCCTTGTGAC AGACTCCGGC
1551 CTGAAGGCAG AATGAACCCA CACCTGGAGT GAGGAAGGGG GCCTGGCACG
1601 GTTGGCCAGG CTCTGCCTGA TTGCCAGCCA GCGGGCATCT GAAGCCGGGT
1651 CCTTCGCCCG CCGGAGGCTG CCGTCCGTCT CTCTGCTGC GCTCGTGCCA
1701 GCTCCGTGGG TGTCCTCCCA GGGAGCTTCT CTCTCAACA GGCCTTGCCA
1751 GGCTGGGGTG AGAGGTGATA GAGGCAGCAC TGTGCATGAT TCCGAGAGGG
1801 TGTGGTGGCA CTGCCAGCCG ACTGCTGACA GCTTGGGAGC TGCTGTGCCC
1851 AGGACGTGGG TTCAGCGTGG GCGAGGAAAG CCTGGCGAGC GTGGCCCTGT
1901 AAAAGCTTTC TGAGGCGGGA GCGCTCACT TACCTGTGAC TGCTGGGGCG
1951 CTGCGTGTAG CATCTTGGCC TACAGGACAG ATTTTAGGTG ACACCTGGTT
2001 ATGACAGTCA GAAATTTGAG AAGCTTCTCA CAAGTGATGC ACTTTAAATA
2051 ATCTGCATGC CATTGAGACA CCTGCATGTC TGGTGTGTGT GGTTCAGGTG
2101 TCTTGCCGCC GGCCTTCGGA TGTAAACCCA CTGATAACGG ACAGAAAGAG
2151 AATGCCCACA AGTGGGTCTT CTGTGGAAGA TGCAGAAAGG GGAAGTTAGT
2201 GCTTACATTT TAGTCTTTTT CTCCCTCAA AAAATAGGTT AAGTTTCACT
2251 GCCAGCTAGA AAATACTGCT TTCTGCCATC GATTGGGGGT GCTTTTGTG
2301 AAATATACGT TTGATAAATA TTTATTTTTG TAACTTGAA GTGTGTGGTG
2351 GCGGTGGGGG AGGACATGCG TGGCAGCAGG CGCCTTCTTC AGCTGTGGGT
2401 CCTTAAAGGCC TTTGATCCTT TGAAGAAGAA AGACATGGTA TTTGTTACGC
2451 AGACGCCGAC CACTCAGACG GAGGGGCCCC TGGGATTCCC TGTCTCAGAT
2501 GGCTGGTCT TACGCTGTG TAGATTTCTT CTCCATTGGG AATGAAGGTG
2551 TCAGGCGGGA CTGGAACGTT CTAGATGGTA TGTCCGTGA TATTAACAAC
2601 TCTAACCAG GACAGACCAC AAGCCCACT CAGAGGCTC ACTGTGCTGG
```

```
2651 GGGCTTCGGT GTCCAGGCGC CCAGGTGTGG CCACCAGCAC CGGTTTCTGC
2701 CTTGCGCTTG CTGGGGTGCA GTGAGACTGC CACACGCGTG CACATGTGGC
2751 TCTGTGGGTG TCTCCTAGAG AGGACGTGGC CCCTGCTGCC AGCCCTTGAG
2801 CAGCCCGTGT GGGGGCCCGA GGGACCCACA CAGTGGGGGC CAGCCTCGCT
2851 GGAGGGAGAG CAACCCCTTG CCGATGACCA CGCTTGCCGC CATCTCTTAG
2901 TTTTCTTTT CACAAGCGCT TTATTTTTT AATAGACAAA TCACATTTTG
2951 CAAGGCCTTT AATTAATAA GATTCTTCTT TCCTTCATTT TATGCTTTAT
3001 TTCCTGTTT AAGGCTTACT GTAGAAGTGG CTTACTGTAG AAGCAGCTTG
3051 CTGAGCCCTT CCGAGCGGTC CCCAGAATTA GCTGGTTCAC AACCCCCACC
3101 CTCCCCCGCC CCGCCTGTG TCAGGTGTGG ATGAGTCTGT CACACTCAGA
3151 AGGACAGGCT TGTCTGCCAG CTCACAAGGG GAGGCTGCAG TGGGTTTGGG
3201 AGCTGGGTTT AGGCCCTGGT TGTCTGAGGG CCCAGGCCCT GCCAGCCTCT
3251 GCTGCTCCTG CTCTGGGTT TGAAGATGCA GGCCGATCGC CAGCTCCGTG
3301 GCAGCGGTCA CTAAGGACAG CTTGACTGTG CCATCTTGGG GCCTCAGGCG
3351 GGGCTCCGGA GATAGAAGAC AGGTCGCCGG AGGCTCCCCC TCCTCTCCTC
3401 TCCCTCTGCG AGATGCTCCC TGGGCGCTAC CTTGCAGGCT GCCAGGCAGG
3451 AGTGGTCTCA GAACGTGCGC TTCTGATTAT TTTACTGGGG TCCATTGTCC
3501 AGATTTTCTT TTGATTGTAA AATATATTTT TACTTTTATG TCTTCTAATT
3551 TAATAAATGA TCCATATAAA AATAGAGAAA TAAAGTCCTT TAAGGGAAGG
3601 TTTAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 112 bp to 933 bp; peptide length: 274
Category: similarity to unknown protein
Classification: no clue

```
1 MATTVSTQRG PVYIGELPQD FLRITPTQQQ RQVQLDAQAA QQLQYGGAVG
51 TVGRLNITVV QAKLAKNYGM TRMDPYCRLR LGYAVYETPT AHNGAKNPRW
101 NKVIHCTVPP GVDSFYLEIF DERAFSMDDR IAWTHITPE SLRQGVKVEDK
151 WYSLSGRQGD DKEGMINLVM SYALLPAAMV MPQPVVLMF TVYQQGVGVV
201 PITGMPAVCS PGMVPVALPP AAVNAQPRCS EEDLKAIQDM FPNMDQEVIR
251 SVLEAQRGNK DAAINSLQGM GEEP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2dl5, frame 1

TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2,
N = 1, Score = 385, P = 1.1e-35

>TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2
Length = 457

HSPs:

Score = 385 (57.8 bits), Expect = 1.1e-35, P = 1.1e-35
Identities = 77/182 (42%), Positives = 118/182 (64%)

```
Query: 4 TVSTQRGPVYIGELPQDFLRIT-PTQQQRQVQLDAQAAQQLQYGGAVGTGRLNITVVQA 62
      TV+ +R V +GELP FLR+ P QQ + ++ Q + + + T GRL+TT+++A
Sbjct: 5 TVAERRRQVLVGELPPHFLRLAVPIQQTAEPEI-VQP-RMVSVFPP-NTRGRLSVTILEA 61

Query: 63 KLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIFDE 122
      L KNYG+ RMDPYCR+R+G ++T A N + P WN+ ++ +P V+S Y++IFDE
Sbjct: 62 NLVKNYGLVRMDPYCRVRVGNVEFDTNVAANAGRAPTWNRTLNAYLPMNVESIYIIFDE 121

Query: 123 RAFSMDDRIAWTHITIPESLRQGVKVEDKWYSLSGRQGGDDKEGMINLVMSYAL--LPAAMV 180
      +AF D+ IAW HI +P ++ G D+++ LSG+QG+ KEGMI+L S+A LP
Sbjct: 122 KAFGPDEVIWAHIMLPLAIFNGONIDEYFQLSGQQGEGKEGMIHLHFSFAPIDLPLQQA 181
```

Query: 181 MPPQP 185

P +P

Sbjct: 182 APAEP 186

Score = 92 (13.8 bits), Expect = 1.8e-01, P = 1.7e-01
Identities = 26/68 (38%), Positives = 38/68 (55%)

Query: 194 QQGVGYVPITGMPAVCSPGMVVPV--ALP--PAAVNAQPRCSEEDLKAIQDMFPMNDQEV 249

QQG G + + +P +P+ A P PA +EED K IQ+MFP +D+EVI

Sbjct: 156 QQGEGKEGMIHLHFSFAPIDLPLQQAAPAEPAAPLPVEITEEDTKEIQEMFPIVDKEVI 215

Query: 250 RSVLEAQR 257

+ +LE +R

Sbjct: 216 KCILEERR 223

Pedant information for DKFZphtes3_2d15, frame 1

Report for DKFZphtes3_2d15.1

[LENGTH] 274
[MW] 30281.97
[pI] 5.68
[HOMOL] TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2 4e-36

[PFAM] C2 domain
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 16.42 %

SEQ MATTVSTQRGPVYIGELPQDFLRITPTQQQRQVQLDAQAAQQLQYGGAVGTGRLNITVV
SEGXXXXXXXXXXXXXXXXX.....
PRD cccccccccceccccccccccccchhhhhhhhhhhhhhhcccccccececeeh

SEQ QAKLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIF
SEG
PRD hhhhhhhccccccccchhhheeeeecccccccccccccecccccccccececeeh

SEQ DERAfSMDDRIAWTHITIPESLRQGVKEDKWYSLSGRQGDDEKGMINLVMSYALLPAAMV
SEGXXXXXXXXX
PRD cccccccccceccccccccccccccccceccccccccccccceceehhhhhhhhhc

SEQ MPPQPVVLMPTVYQQGVGYVPITGMPAVCSPGMVVPVALPPAAVNAQPRCSEEDLKAIQDM
SEGXXXXXXXXX.....
PRD cccccceccccccccccccccccceccccccccccccceccccchhhhhhhhhc

SEQ FPNMDQEVIRSVLEAQRGNKDAAINSLQMGEEP
SEG
PRD cccccchhhhhhhhhccccchhhhhhhhhcc

(No Prosite data available for DKFZphtes3_2d15.1)

Pfam for DKFZphtes3_2d15.1

HMM_NAME C2 domain
HMM *LTVrIIeARNLWkMDMnGfSDPYVKVdMdPdpkDtkKWktTiWNNGLN
L+++++A+ + + M+ DPY+++ + + +T T +N N
Query 55 LNITVVQAKLAKNYGMT-RMDPYCRLRLGYAVY----ETPTAHNGAKN 97
HMM PVWNEEeFvFedIPyPdlqrkMLRFaVWDWDRFSRBDFIGHCi*
P+WN + +P + + ++++D+ FS +D I+ +
Query 98 PRWN-KVIHCT-VPPGVDSF---YLEIFDERAFSMDDRIAWTH 135

DKFZphtes3_2e12

group: Transcription Factors

DKFZphtes3_2e12 encodes a novel 849 amino acid protein with similarity to Zinc finger proteins.

The new protein is a putative transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein, which is only found in cytochrom C related proteins.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to finger proteins

complete cDNA, complete cds, 5 EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3205 bp

Poly A stretch at pos. 3192, polyadenylation signal at pos. 3171

```
1  GGCACGGCCG GGTCTGGCT GGCCAAACGA GGCTCGCGGA AGCAGCAGCC
51  GCCCCCTGAC CGCAGCTGGA TTTGAAGAT TGATCCAAGG GACTGTATTA
101 ATTTTCAGGAA TTGATTGAA AGACACTGGC TCTGCCACTT AACAGCCATG
151 TAACCTTGGA TATGGAAGAA AGTAGCAGTG TTGCCATGTT GGTGCCAGAT
201 ATTGGGGAAC AGGAAGCTAT ACTGACTGCT GAAAGTATCA TCAGTCCTTC
251 ATTGGAATTT GATGAACAAA GAAAACTAA ACCAGATCCA TTAATCCATG
301 TTATCCAGAA GTTAAGCAAG ATAGAAAAAT GAAAAGTCAC AAAAATGTCT
351 TTTAATTGGG AAGAAACGCC CACGTTCAGT TGCTGCAACA CACTCTCTTG
401 AAACCCAAGA ACTTTGTGAG ATTCCGGCTA AAGTAATCCA GTCACCTGCT
451 GCTGATACTA GAAGGGCTGA GATGTCACAA ACAAAATTTA CCCCTGACAC
501 TCTTGCCAGC AATGAAGGGA AGGCTATGTC TTATCAGTGT AGCCTTTGTA
551 AGTTTCTATC ATCATCCTTT TCCGTGTTAA AAGATCATAT TAAGCAACAT
601 GGTCAGCAAA ATGAAGTGAT ACTGATGTGC TCAGAGTGCC ATATTACATC
651 TAGAAGCCAG GAGGAACCTT AAGCCACGCT GGTGAATGAC CATGACAATG
701 ATGCCAATAT CCACACCCAA TCCAAAGCCC AACAGTGCGT AAGCCCCCTC
751 AGCTCTTTGT GTCGAAAAC CACAGAAAGA AATGAAACCA TTCCAGATAT
801 CCCAGTAAGT GTGGACAATC TACAGACTCA TACTGTCCAA ACTGCATCTG
851 TGGCAGAAAT GGTAGGAGG AAATGGTATG CATACGAACA GTACGGCATG
901 TATCGATGCT TGTTTGTAG TTATACTTGT GGCCAGCAGA GAATGTTGAA
951 AACACACGCT TGGAAACATG CTGGGGAGGT TGATTGCTCC TATCCAATCT
1001 TTGAAAATGA AAATGAACCC CTAGGCCCTG TGGATTCTTC AGCAGCTGCT
1051 GCGCCTGGTG GGTTCGATGC AGTCGTCTAT GCTATTGGAG AGAGTGAACT
1101 GAGTATCCAC AATGGGCCAT CAGTGCAAGT GCAGATTGCG AGCTCAGAAC
1151 AGTTATCATC TTCATCTCCT TTAGAACAGA GTGCAGAAAG AGGAGTACAC
1201 CTAAGTCAGT CAGTTACCTT GGACCCCAAT GAGGAAGAAA TGCTAGAAGT
1251 GATTTCTGAT GCAGAGGAGA ATCTGATTCC TGATAGCCTG CTTACATCAG
1301 CACAGAAAAT CATCAGCAGC AGCCCCAATA AAAAAGGGCA TGTTAACGTG
1351 ATAGTGGAGC GATTGCCAAG TGCTGAAGAA ACCCTTTCAC AGAAGCGCTT
1401 CCTCATGAAC ACTGAAATGG AAGAAGGGAA GGACCTGAGC CTGACAGAAG
1451 CTCAGATTGG GCGCGAAGGA ATGGATGATG TTTATCGTGC TGATAAATGT
1501 ACTGTTGATA TTGGGGGATT GATCATAGGC TGGAGCAGTT CAGAGAAAAA
1551 AGACGAGTTA ATGAATAAAG GCCTGGCTAC TGATGAGAAT GCCCACCAG
1601 GCCGGAGAAG GACAAATTCT GAGTCTCTTC GATTACACTC ATTAGCTGCA
1651 GAAGCCCTTG TCACAATGCC TATAAGAGCT GCAGAGTTGA CAAGAGCCAA
1701 CCTGGGGCAC TATGGAGATA TAAACCTTTT AGATCCAGAT ACTAGTCAAA
1751 GGCAAGTAGA TAGTACATTG GCAGCGTACT CAAAAATGAT GTCGCCACTT
1801 AAAAATCTTT CAGATGGATT AACTAGTCTT AACCAAAGCA ACTCCACCTT
1851 GGTAGCACTC CCAGAGGGTA GGCAGGAATT GTCAGATGGG CAGGTTAAGA
1901 CAGGCATCAG CATGTCCTTA CTCACCGTCA TTGAAAAATT GAGAGRAAGG
1951 ACAGACCAAA ACGCTTCAGA CGATGACATT TTGAAAGAGT TGCAGGACAA
2001 CCCCCAGTGC CAACCAACAA GCGATACAAG TTTGTCGGGA AACAATGTGG
2051 TGGAAATACAT CCCGAATGCT GAACGACCTT ACCGTTGCCG CCGTGTTCAC
2101 TACACAAGTG GCAACAAGGG CTACATCAAG CAGCACTTAC GAGTCCATCG
2151 ACAGAGACAG CTTATCAGT GTCCTATCTG CGAGCACATA GCGGACAACA
2201 GCAAAGATTT GGAGAGTCAC ATGATCCACC ACTGTAAGAC AAGAATATAC
2251 CAGTGCAAGC AGTGTGAAGA ATCCTTCCAT TATAAGAGTC AATTGAGGAA
2301 CCATGAGAGA GAACAGCACA GTCTTCCAGA TACCTTGTCA ATAGCAACTT
2351 CTAATGAGCC AAGAATTTCC AGTGATACAG CTGATGGAAA ATGTGTCAG
2401 GAAGGGAATA AGTCTTCAGT CCAGAAACAA TATAGATGTG ATGTGTGTGA
2451 TTATACAAGT ACAACATATG TTGGTGTGAG AAACCACAGG CGAATCCATA
2501 ACTCTGATAA GCCGTACAGA TGCTCTCTGT GTGGGTATGT GTGTAGCCAT
2551 CCTCCTTCTT TGAAGTCTCA TATGTGGAAG CATGCAAGTG ACCAAAAATTA
```

```

2601 CAACTACGAA CAAGTAAACA AGGCTATTAA CGACGCGATT TCACAAAGTG
2651 GCAGAGTTCT GGGGAAATCC CCTGGAAAGA CTCAATTAAA GAGCAGTGAA
2701 GAGAGTGCAG ATCCCGTCAC TGGAACTTCG GAAATGCAG TGTCTCTTC
2751 AGAACTGATG TCCCAGACTC CCAGTGAAGT TCTGGGTACC AACGAGAATG
2801 AGAACTGAG CCCTACAAGT AATACCTCAT ATAGTTTAGA AAAAACTCC
2851 AGTCTGGCCC CTCCTAGCAT GGAGTACTGC GTTTTACTCT TCTGCTGTG
2901 TATTTGTGGT TTGGAATCAA CCAGCAAAGA AAACCTCTTG GATCATATGA
2951 AAGAGCACGA GGGTGAAATT GTAAACATCA TCCTGAATAA GGACCACAAT
3001 ACAGCTCTAA ACACAAATTA GGTGGAATAA TGACTCGAGC AGGAAAGCAG
3051 TAGAAGAGGA TTCCTTCACC ACAGTTTACC CTTTACGCTG TCAGACAACT
3101 TCCTGCCACA GAAGAAGTCG TTGATGTGAT TTTTGAGGAA ATGACAGATG
3151 TGACTTTGGA ACCAAACTTG TAATAAAGG AATTCCAAAT GGAACAAAAA
3201 AAAAA

```

BLAST Results

No BLAST result

Medline entries

90301500:
Cloning and sequencing of a zinc finger cDNA expressed in mouse testis.

92310982:
zfp-37, a new murine zinc finger encoding gene, is expressed in a developmentally regulated pattern in the male germ line.

Peptide information for frame 1

ORF from 472 bp to 3018 bp; peptide length: 849
Category: similarity to known protein

```

1 MSQTNFTPDT LAQNEGKAMS YQCSLCKFLS SSFVSLKDHI KQHGOQNEVI
51 LMCSECHITS RSQEELEAHV VNDHDNDANI HTQSKAQQCQ SPSSSLCRKT
101 TERNETIPDI PVSVDNLQTH TVQTASVAEM GRRKWAYAYE YGMYRCLFCS
151 YTCGQQRMLK THAWKHAGEV DCSYPIFENE NEPLGLLDSS AAAAPGGVDA
201 VVIAIGESSEL SIHNGPSVQV QICSSEQLSS SSPLEQSAER GVHLSQSRTL
251 DPNEEEMLEV ISDAEENLIP DSLTSAQKI ISSSPNKKGH VNVIVERLPS
301 AEETLSQKRF LMNTEEMEGK DLSLTEAQIG REGMDDVYRA DKCTVDIGGL
351 IIGWSSSEKK DELMNKGLAT DENAPPGRRR TNSESLRLHS LAEALVTMP
401 IRAAELTRAN LGHYGDINLL DPDTSORQVD STLAAYSKMM SPLKNSDGL
451 TSLNQSNSTL VALPEGRQEL SDGOVKTGIS MSLLTVIEKL RERTDQNASD
501 DDILKELQDN AQCPQNSDTS LSGNNVVEYI PNAERPYYRCR LCHYTSGNKG
551 YIKQHLRVHR QRQPYQCPIC EHIADNSKDL ESHMIHCKT RIYQCKQCEE
601 SFHYKQLRN HEREQHSPLD TSLIATSNEP RISSDTADGK CVQEGNKSSV
651 QKQYRCDVCD YTSTTYVGVN NHRRIHNSDK PYRCSLCGYV CSHPPSLKSH
701 MWKHASQNY NYEQVNKAIN DAISQSGRVL GKSPGKTQLK SSEESADPVT
751 GSSENAVSSS ELMSQTPSEV LGTNEKLS PTSNTSYSLE KISSLAPPSM
801 EYCVLLFCCC ICGFESTSKE NLLDHMKHE GEIVNIILNK DHNTALNTN

```

BLASTP hits

Entry S10245 from database PIR:
finger protein, testis - mouse
Score = 265, P = 8.4e-23, identities = 61/205, positives = 91/205

Entry S22954 from database PIR:
finger protein zfp-37 - mouse
Score = 265, P = 9.1e-22, identities = 61/205, positives = 91/205

Entry AF031657.1 from database TREMBL:
gene: "zfp94"; product: "zinc-finger protein 94"; Rattus norvegicus
zinc-finger protein 94 (zfp94) gene, partial cds.
Score = 243, P = 1.6e-21, identities = 57/190, positives = 85/190

Alert BLASTP hits for DKF2phtes3_2e12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2el2, frame 1

Report for DKFZphtes3_2el2.1

[LENGTH] 849
[MW] 94325.42
[pI] 5.47
[HOMOL] PIR:A54661 zinc finger protein 2NF41 - human (fragment) 2e-22
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YJL056c] 3e-09
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YJL056c] 3e-09
[FUNCAT] 04.03.01 trna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT] 04.01.01 rrna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YOR113w] 4e-07
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YGL209w] 2e-04
[FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YNL027w] 2e-04
[FUNCAT] 11.01 stress response [S. cerevisiae, YMR037c] 3e-04
[BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins
[SCOP] dlmeyg_9.6.1.1.1 a designed zinc finger protein [syntheti 8e-06
[PIRKW] nucleus 8e-18
[PIRKW] RNA binding 5e-13
[PIRKW] duplication 7e-13
[PIRKW] tandem repeat 1e-21
[PIRKW] spermatogenesis 6e-16
[PIRKW] zinc 9e-21
[PIRKW] zinc finger 1e-21
[PIRKW] DNA binding 1e-21
[PIRKW] metal binding 3e-15
[PIRKW] phosphoprotein 5e-13
[PIRKW] leucine zipper 1e-13
[PIRKW] alternative splicing 6e-18
[PIRKW] eye lens 2e-16
[PIRKW] oocyte 1e-12
[PIRKW] transcription factor 6e-18
[PIRKW] segmentation 7e-13
[PIRKW] embryo 1e-12
[PIRKW] transcription regulation 2e-19
[PIRKW] homeobox 2e-08
[SUPFAM] POZ domain homology 7e-15
[SUPFAM] transcription factor Krueppel 7e-13
[SUPFAM] zinc finger protein ZFP-36 1e-21
[SUPFAM] homeobox homology 2e-08
[SUPFAM] unassigned homeobox proteins 2e-08
[PROSITE] CYTOCHROME_C 1
[PROSITE] MYRISTYL 10
[PROSITE] ZINC_FINGER_C2H2 3
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 18
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 7
[PFAM] Zinc finger, C2H2 type
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 5.65 %

SEQ MSQTNFTPDTLAQNEGKAMSYQCSLCKFLSSSFVLKDHQKHGQQNEVILMCSECHITS
SEGXXXXXXXXXXXXXXXXX
lmeyF
SEQ RSQEELAHVVDNDNDANIHTQSKAQCVSPSSSLCRKTTTERNETIPDIPVSVNQLQTH
SEG
lmeyF
SEQ TVQTASVAEMGRKWKYAYEQYGMRYCLFCSYTCGQQRMLKTHAWKHAGEVDCSYPIFENE
SEG
lmeyF
SEQ NEPLGLLDSSAAAAAPGGVDAVVIAIGESLSIHNGPSVQVQICSSEQLSSSSPLEQSAER
SEGXXXXXXXXXXXXXXXXX
lmeyF
SEQ GVHLSQSVTLDPNEEMLEVISDAEENLIPDSLTSQAQKIISSSPNKKGHVNVIVERLPS
SEG
lmeyF

```

SEQ      AEETLSQKRFMNTMEEGKDLSTEAQIGREGMDDVYRADKCTVDIGGLIIGWSSEKK
SEG      .....
lmeYF    .....

SEQ      DELMNKGLATDENAPPGRRTNSESRLHSLAAEALVTMPIRAAELTRANLGHYGDINLL
SEG      .....
lmeYF    .....

SEQ      DPDTSQRQVDSTLAAYSKMMSPLKNSSDGLTSLNQSNTLVALPEGRQELSDGQVKTGIS
SEG      .....
lmeYF    .....

SEQ      MSLLTVIEKLRERTDQNASDDDLKELQDNAQCQPNSDTSLSGNNVVEYIPNAERP YRCR
SEG      .....
lmeYF    .....TTTEETT

SEQ      LCHYTSGNKGYIKQHLRVHRQRQPYQCPICEHIADNSKDLESHMIHCKTRIYQCKQCEE
SEG      .....
lmeYF    TTTCEETTHHHHHHHHHHTTCCCEETTTTTEECCHHHHHHHHHHHCCCCCEETTTTE

SEQ      SFHYKSQLRNHEREQHSLPDTLSIATSNEPRISSDTADGKCVQEGNKSSVQKQYRCDVCD
SEG      .....
lmeYF    EECCHHHHHHHHHHC.....

SEQ      YTSTTYVGRNHRRIHNSDKPYRCSLCGYVCSPPSLKSHMKHASDQNYNEQVNKAIN
SEG      .....
lmeYF    .....

SEQ      DAISQSGRVLGKSPGKTQLKSSEESADPVTGSSSENAVSSSELSQTPSEVLGTNENEKLS
SEG      .....
lmeYF    .....

SEQ      PTSNTSYSLEKISSLAPPSMEYCVLLFCCCICGFESTSKENLLDHMKHEGEIVNIILNK
SEG      .....
lmeYF    .....

SEQ      DHNTALNTN
SEG      .....
lmeYF    .....

```

Prosites for DKFZphtes3_2e12.1

PS000001	104->108	ASN_GLYCOSYLATION	PDOC00001
PS000001	445->449	ASN_GLYCOSYLATION	PDOC00001
PS000001	454->458	ASN_GLYCOSYLATION	PDOC00001
PS000001	457->461	ASN_GLYCOSYLATION	PDOC00001
PS000001	497->501	ASN_GLYCOSYLATION	PDOC00001
PS000001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS000001	784->788	ASN_GLYCOSYLATION	PDOC00001
PS000004	98->102	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS000005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS000005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS000005	357->360	PKC_PHOSPHO_SITE	PDOC00005
PS000005	385->388	PKC_PHOSPHO_SITE	PDOC00005
PS000005	425->428	PKC_PHOSPHO_SITE	PDOC00005
PS000005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS000005	696->699	PKC_PHOSPHO_SITE	PDOC00005
PS000005	726->729	PKC_PHOSPHO_SITE	PDOC00005
PS000005	817->820	PKC_PHOSPHO_SITE	PDOC00005
PS000006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS000006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS000006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS000006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS000006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS000006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS000006	314->318	CK2_PHOSPHO_SITE	PDOC00006
PS000006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS000006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS000006	381->385	CK2_PHOSPHO_SITE	PDOC00006
PS000006	485->489	CK2_PHOSPHO_SITE	PDOC00006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS000006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS000006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS000006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS000006	758->762	CK2_PHOSPHO_SITE	PDOC00006
PS000006	766->770	CK2_PHOSPHO_SITE	PDOC00006
PS000006	817->821	CK2_PHOSPHO_SITE	PDOC00006

PS00007	331->339	TYR_PHOSPHO_SITE	PDOC00007
PS00007	703->711	TYR_PHOSPHO_SITE	PDOC00007
PS00007	596->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	142->148	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	241->247	MYRISTYL	PDOC00008
PS00008	349->355	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	478->484	MYRISTYL	PDOC00008
PS00008	645->651	MYRISTYL	PDOC00008
PS00008	751->757	MYRISTYL	PDOC00008
PS00008	772->778	MYRISTYL	PDOC00008
PS00009	130->134	AMIDATION	PDOC00009
PS00009	376->380	AMIDATION	PDOC00009
PS00028	146->167	ZINC_FINGER_C2H2	PDOC00028
PS00028	684->705	ZINC_FINGER_C2H2	PDOC00028
PS00028	595->617	ZINC_FINGER_C2H2	PDOC00028
PS00190	53->59	CYTCHROME_C	PDOC00169

Pfam for DKFzptes3_2e12.1

HMM_NAME Zinc finger, C2H2 type

HMM *CpwPDCgKtFrrwsNlrRHMRT.H*
C++ C+ T R+++L++H H

Query 53 CSE--CHITSRSQEELAHVVN-DH 74

23.25 (bits) f: 539 t: 559 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
Query *CpwPDCgKtFrrwsNlrRHMRT.H*
C C++T ++ ++H+R+H
dkfzptes3 539 CRL--CHYTSNGYIKQHLRVH 559

Query f: 567 t: 587 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
HMM *CpwPDCgKtFrrwsNlrRHMRT.H*
CP+ C+ ++ +L+ HM+ H
Query 567 CPI--CEHIADNSKDLESHMIHH 587

33.47 (bits) f: 595 t: 616 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
Query *CpwPDCgKtFrrwsNlrRHMRT.H*
C+ C+++F ++S+LR+H R H
dkfzptes3 595 CKQ--CEESFHYKSQLRNHERE-QH 616

Query f: 656 t: 676 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
HMM *CpwPDCgKtFrrwsNlrRHMRT.H*
C++ C++T ++ R+H+R+H
Query 656 CDV--CDYTSTTYVGVNRHRIH 676

24.53 (bits) f: 684 t: 704 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
Query *CpwPDCgKtFrrwsNlrRHMRT.H*
C+ CG++ +++ +L+ HM H
dkfzptes3 684 CSL--CGYVCSHPPLKSHMWKH 704

Query f: 809 t: 829 Target: dkfzptes3_2e12.1 similarity to finger proteins
Alignment to HMM consensus:
HMM *CpwPDCgKtFrrwsNlrRHMRT.H*
C+ CG ++++NL HM+ H
Query 809 CCI--CGFESTSKENLLDHMKEH 829

DKFZphtes3_2f14

group: testes derived

DKFZphtes3_2f14 encodes a novel 129 amino acid protein with very weak similarity to human omega protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to omega protein

complete cDNA, complete cds, 1 EST hit

Sequenced by EMBL

Locus: unknown

Insert length: 2353 bp
Poly A stretch at pos. 2341, no polyadenylation signal found

```
1 GCAGATTCTC CAGGCCAGC ATCTGCCTCA CCGTGGCCCC CCACAAGCCA
51 AGCGCCTGCC TTTCAGCAGC CTCTACACAC CCAGCTCCTG CCACCCAATG
101 GCTCTTTAGG CCAAGCTCAT ACCTCAGCAT GATTTTCCA GGCCCAACTT
151 TTGTCTCATG GCAACCTTCC CTGGCCAAGT TTCCACCTAT TTCCTGGCAG
201 CCTGGACAGG CCCAGGTCTT GCCACACACT GGCTCTCTTA CGCCAGCTTC
251 ATGCCTCACA GTGGCCTCTC CAGGCCAGC TCCTGTCCCG GGACATCATC
301 TCCAGGCCCA AACTTCTCTC AAGTCGGCCT CTCCAGGCC AGTTGCTGCC
351 TCCCGGCATT CTCTCCAGGC CTAGCTCTTC CTCTGGCTG TATCTACAAG
401 ACCAACTCCT GCCTCACAA ACCTTTTAT GGCTCAGCTC CTGCCCAACT
451 ACTGCCGGCC TTGTAGGCC CAAAACCTCC TCAAGTCAAG CTCTTAGGCG
501 CCACCTTCTG CTTGCAGTG GCCTGTACAG ACCCAGCTCT GGCTTGAGAA
551 CAGCCTCTGC AGGCCCTGCT CTTGCCTCTT AGCTCCCTCT CCAGGCCCAT
601 CTCTTGCCCTC ACAGTGGCTT CCGTGGGCCA AGTCCCGCC TGCCTCCAG
651 CAGCCTCAAC AGGCCCTAGCT CCTCCCTCAC AATGGCTTGT TTAGGTCAG
701 TTGATGCCTC TGGCAACCTG TCCAGGCCCA GCTCCTGCCT CACACTGGCC
751 TCTCTAGGCC GAGGTCCTTT CTCATACTGG CCGTTTATAG CCCAGCTCAT
801 TCCTCTTGTC ATCTCTCCAG GCCCAGCTTT TGCTGTGTTG TGGCCTCTAC
851 CTCACAGTGC ACCTTCCAGT CCCACCTCTT GCCTCACCAT GGCCTCTCTT
901 GACCAAGTTC CTGCCTTTCC GCAGCCTCTA CAGGCCTAGC TGCTGCCTCC
951 CAATGGCCTT TGTAGGCCAC GCTCATGCCT CACTGTGGCC TTTCCAGGCC
1001 TAGCTTTTGC TTTTGGCCA CTCCAGGCC AGAACTTCCC CCAGTCAGCC
1051 TCTCCAGGCC CAGCTCTTCC TCCAGCAAC CTCTGCAGGC CCAATCATC
1101 CTCAAATTGG CCTCTTCTT CCCAGCTCCT GCCTCCTGGT GGCCTCTGAA
1151 GACCCAAATC GTCTCCAGT TGGTTTTC AGGCCAGCT CCTGCCTTTT
1201 GGTGGCCTCT CCAGGTGCAA AACTTCTTCC CATCAGCCTG TCCAGGCCCA
1251 GCTCATGCCT CTGGTGGCC TTCTCAGGCC CTGCTTTTGA CTGGTGGGCC
1301 TCTTCAGGCC CAGAACTTGA ACTCAAGTCA GCCTCTCCAG GCCCAGTCC
1351 TGCTTCTTGA AGGTCTGTAC AGGCCAGCC TCTACCTCAC AGCGACTCT
1401 CCACACCCAG CTCTTGCCCTC ACTGTAGCCT CCCAGTCCA AAACCTCTGC
1451 CTTTTGGCAG CTTCGACAAG CCCAGCTCCT GCCTTTCAAT GACCTCTTGA
1501 GGCCCGCCTC ATTCCTTACA ACGGCCTTTC CAGGCCAGT TTTTCCCTTT
1551 TGGCGGCCTC TCCAGGCCCA GAACTTCTC AAGTCGGCCT CTTTAGGCC
1601 AGTTGCTGCC TCCTGGCCTC CTCTGCAGGC CGAGCTCTTC CTCCCTGCTG
1651 TGTCTACAGG CCAACTCCT GCCTCACAA AACCTCCTTG GACTCAGCTT
1701 CTGCCAGCTC CTGTGTGGCC TTTGTAGGCT CAAATTTTC TCAAATCAAG
1751 CTCTCCAGGC CTACTGTGAG CCTCGTGCCA GCCTAAACAG GCCCAGCTCC
1801 TGCTTGACAA TGGCCTCTCC AGGCTTTTCT CCGCCTCGC AGCAGGCTTT
1851 CCAGGCCAG CTCTTGCCCTC ATGGTGGCCT TCCCGGCCA TGTTCCTATC
1901 TGACTTCTGG CAGCCTCAAC CGGCCAGCT TCTGCCTCAC ACTGGCCTCT
1951 CTAGGCCAG CTCTTTTTC ACAGTGGCCT CACTACGCC ATCTCCTAGC
2001 TCAGATCTGC CTCCAAGAC CCAGCTCCTG TCTCATGGTG GTCTCTCTTA
2051 CACCAGCTCC TGCTCACAA TGGCCTCGTC TGGCCATCT TCTGCCTCAC
2101 AGTGGCCACT CAAGGCCCAT CTTTGGCCTC ATGGTAGCCT CTCTGGTTT
2151 TGCTCTTGCC TCACAGTTGC CTCTTCCAGA TCCAGCTTTA AGCCTTTGAT
2201 GGTCAACAGC ATCAAGGAGC CTAAGCTTC CTGGACTCT CATTTGTTCA
2251 CTTTACAGCA GAGTGCCTTA GCAAAAAC TGCTCTAAC TTGAGAGTGG
2301 ATTTCTGACA AATCGATAGT AAATCTGCC TGTGTGGTTT CAAAAA
2351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 158 bp to 544 bp; peptide length: 129
Category: similarity to known protein

1 MATFPGQVST YFLAAWTGPG PATHWPLYAQ LMPHSGLSRP SSCPGTSSPG
51 PKLPQVGLSR PSCCLPAFSP GLALPPGCIY KTNSCLTTTF YGSAPAQLLP
101 AFVGPKLPQV KLFRTFCCLA VACTDPALA

BLASTP hits

Entry I70697 from database PIR:
omega protein - human (fragment)
Score = 79, P = 2.8e-03, identities = 32/94, positives = 38/94

Alert BLASTP hits for DKFZphtes3_2f14, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2f14, frame 2

Report for DKFZphtes3_2f14.2

[LENGTH] 129
[MW] 13421.76
[pI] 9.14
[PROSITE] MYRISTYL 2
[KW] Irregular
[KW] LOW_COMPLEXITY 10.85 %

SEQ MATFPGQVSTYFLAAWTGPGPATHWPLYAQLMPHSGLSRPSGPCGTSSPGPKLPQVGLSR
SEGXXXXXXXXXXXXXXXXX.....
PRD cccccccceehhhhhcc

SEQ PSCCLPAFSPGLALPPGCIYKTNSCLTTTFYGSAPAQLLPFVGPKLPQVKLFRTFCCLA
SEG
PRD cccccccccccccccccccccceeccccccccccccccccccccccccccccccccc

SEQ VACTDPALA
SEG
PRD cccccccc

Prosite for DKFZphtes3_2f14.2

PS00008 6->12 MYRISTYL PDOC00008
PS00008 92->98 MYRISTYL PDOC00008

(No Pfam data available for DKFZphtes3_2f14.2)

DKF2phtes3_2g7

group: testes derived

DKF2phtes3_2g7 encodes a novel 359 amino acid protein with similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

complete cDNA, complete cds, 6 EST hits (5 hits are out of a testis library)

Sequenced by EMBL

Locus: unknown

Insert length: 1613 bp

Poly A stretch at pos. 1595, polyadenylation signal at pos. 1557

```
1 GCCACACAGG CTCCTTGGAG TAAGAGTGTG AGAAACTGGA TGAAGACAGC
51 TGTATTCTTT TGAAGCGGTT CGAGATTGGT CTGTCTCTAC CAACTAAAAA
101 CTTCTAGCTT AAGTGCAGAG ATTTAAGGAG ATCAACAAAA ACTCAGTCTA
151 GACATATTAT GAGGCTGGGA GGGTATCAAC AGACTTGAGT TCTTGTCAGC
201 AAGATCACCT GCTTTAATA TTGTCTCAG GGTCTGAGCA CATCTGGAAG
251 TGAGGTCAAT CAAGTTAGAC CCCAAAAACT TTTGTGACAA CAGTGAAGAG
301 GGGAAAAATA ACACACCACA AACATGAACC TCAACCCCCC GACATCTGCT
351 CTTCAGATCG AGGGCAAAGG CAGCCATATT ATGGCTAGAA ATGTAAGCTG
401 CTTCTAGTC AGGCACACCC CTCATCCCAG AAGAGTCTGC CACATCAAAG
451 GCTTGAATAA CATTCCAATC TGTACTGTGA ATGATGATGA GAATGCATTT
501 GGAACATTGT GGAAGTTGG CCAGTCTAAC TACTTAGAGA AGAACAGGAT
551 ACCATTTGCC AATTGCAGTT ACCCCCCGAG CACTGCAGTC CAGAAGAGCC
601 CTGTAAGAGG AATGTCGCCA GCCCCAAACG GTGCCAAAGT GCCTCCACGG
651 CCTCATTCTG AGCCCAGTAG AAAAATTAAA GAGTGTCTCA AAACCTCCAG
701 TGAGAATCCC TTAGTAATTA AAAAGGAAGA AATTAAGGCC AAAAGACCAC
751 CATCACCTCC AAAGGCATGC TCTACTCCTG GCTCCTGTTC TTCAGGGATG
801 ACAAGTACCA AGAATGATGT GAAAGCAAAC ACCATTGTGA TACCAAACCTA
851 TCTGGATCAG GAAATAAAAA TCCTGGCAAA GCTCTGTAGC ATTTTGCATA
901 CTGATTCTCT GGCAGAAGTT TTACAGTGGC TGCTTCATGC AACTTCAAAA
951 GAAAAAGAGT GGGTCTCAGC TTTGATTTCAT TCTGAGCTTG CCGAGATAAA
1001 CCTGTAACT CATCACAGAA GAAACACCTC AATGGAACCA GCAGCAGAGA
1051 CTGGGAAGCC ACCCACAGTT AAATCACCAC CCACAGTTAA ATTGCCCCCA
1101 AATTTTACTG CAAAATCAAA AGTGCTGACC AGAGATACAG AAGGGGATCA
1151 ACCAACCAGA GTGTCAAGTC AAGGATCTGA AGAAAACAAG GAAGTACCAA
1201 AAGAGGCTGA GCACAAGCCT CCACTACTTA TAAGAAGAAA TAATATGAAA
1251 ATACCTGTTG CAGAATATT CAGCAACCA AATTCTCCTC CCAGGCCTAA
1301 CACTCAGGAG AGTGGATCAG CAAAACCACT GTCAGCAAGG AGTATACAAG
1351 AATACAACCT CTGTCCCCAA AGAGCATGTT ATCCTTCAAC ACACCGGAGG
1401 TAGAAGTTCT AGACTGGGTG AATTCTTTCA TGAATATGAG CTTACATTTT
1451 ACATCATCAA ATTATTTTTC AAATGAATAT TTTTGTATT GAGGAATCAA
1501 GTGGTCTCTT TTATGGTGGC ACATGTAAT CTAAAAATAC CTGTATGTAA
1551 TGCTACAAAT AAATATTACT GGAATGATA TTTCCATTG TAGTTAAAAA
1601 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 324 bp to 1400 bp; peptide length: 359
Category: similarity to known protein

BLASTP hits

794

PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	262->265	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	356->359	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	207->211	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00006	272->276	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	158->164	MYRISTYL	PDOC00008
PS00008	284->290	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphtes3_2g7.3)

DKFZphtes3_2h1

group: transmembrane protein

DKFZphtes3 2h1 encodes a novel 116 amino acid protein with weak similarity to C. elegans cosmid C13F10.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to C.elegans C13F10.5

TRANSMEMBRANE 1

Sequenced by EMBL

Locus: /map="2"

Insert length: 1156 bp

Poly A stretch at pos. 1143, polyadenylation signal at pos. 1121

```
1 GGCCATCAAA ATAACATAAC CATGTCATTT GGAGCAACAA AGCCACTGCG
51 GCCTCCATTT GGGCCAAGCT CTGACTGCAA TGATGCCTCT GCCCGACCC
101 GGGCCTCGCT GTGACTGACA ATGCCGCTGC ATCTTTTCAG CAGTCATTGA
151 TGAGGAAGTA TCTACATCCT CCTTCCCACT ACCAGATTTT GCTTGGAGAA
201 AAGCAGTTTC CTGAAATAAT TCTGTGACGA GCTTCTTCCA CATTAGGACA
251 AAAATGCTGG AAGCGGCTCA GCCCAGGGC AGCACATCAG AGACACCATG
301 GAACACAGCC ATTCCTCTGC CGTCGTGCTG GGACCACTCT TTCCTGACCA
351 ATATCACCTT CTGGAAGGTT CTCTCTGGT TGGTCCTGCT GGGACTGTTT
401 GTGGAACCTG AATTGGCCCT GGCATATTTT GTCCTGCTCT TGTCTATTG
451 GATGTACGTC GGGACACGAG GCCCTGAAGA GAAGAAAGAG GGAGAGAAGA
501 GCGCCTACTC TGTGTTCAAT CCAGGCTGTG AAGCCATCCA GGGCACCCCTG
551 ACTGCAGAGC AGTTGGAGCG CGAGTTACAG TTGAGACCCC TGGCAGGGAG
601 ATAGGACCCA GCTGTGCTGT CATGCAGCTA ACCTCTGATG TGGTCTTCCT
651 CACCATTTGG TATGGATTGT ATTCAGGTG TATAGGACTA AGGGCAGCTT
701 GCGGGTTAGC TCTGTGACTG CATAGTTTTT CTACCTTCTT TCCCTGATCT
751 TTTGCTGCCA TTTGATCTTT GATAGTTTTG GTGAAACTCT CTAATAATACA
801 TTCCTGTGG GTCCGACGCA ATTTATAAAA ATTATGTACT CAAGAAGGGA
851 GACCTGTTTG TTTCAATTTCT CATCTGTTTG GGAGATGATT TTAGAGCACT
901 AGAAAGGCAC TGGGGAGATT CTCAGCTTAA AACATCCAGC AGTTTGAAGT
951 ATGATTAGGT ACATCAGGGC TGCATTGTCA ATGTCTCTCT TAAGTCTTTT
1001 AACATTATTA GCAATTTTTT TTTTCCCGGA GAGTTTAGGT TGCAAGTTTT
1051 GGGTTTCTTG TTTGTTTTTG TTTTGCTTCC TGCTTTAATT CTTTAATTTT
1101 CAGTCATTAC TGGTATTGAA AAATAAAATA TCTTAAACAC ATCAAAAAAA
1151 AAAAAA
```

BLAST Results

Entry HS313307 from database EMBL:

human STS SHGC-16715.

Score = 1222, P = 1.4e-48, identities = 248/251

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 254 bp to 601 bp; peptide length: 116

Category: similarity to unknown protein

```
1 MLEAAQPGS TSETPWNTAI PLPSCWDQSF LTNITFLKVL LWLVLLGLFV
51 ELEFGLAYFV LSLFYWMYVG TRGPEEKKEG EKSAYSVFNP GCEAIQGLT
101 AEQLERELQL RPLAGR
```

No BLASTP hits available

TREMBL:CEUC13F10_2 gene: "C13F10.5"; *Caenorhabditis elegans* cosmid
C13F10., N = 1, Score = 141, P = 8.2e-10

HSPs:

Query: 87 VFNPGCEAIQGTLTAEQLEREL 108
VFN CE + G++TAE ER++
Sbjct: 145 VFNDNCERLAGSMTAEHFERDM 166

Report for DKFZphtes3_2h1.2

```

SEQ      MLEAAQPGQSTSETPWNTAIPLPSCDQDSFLTNTIFLKVLVLVLLGLFVELEFGLAYFV
PRD      .....xxxxxxxxxxxxxxxxxxxxxxxx.....
SEQ      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhchhhhh
PRD      .....MMMMMMMMMMMMMMMM.....
MEM
SEQ      LSLFYWMYVGTGRPEEKKEGEKSAYSVFNPGECAIQTGLTAEQLERELQLRPLAGR
PRD      .....xxxxxxxxxxxxxxxxxxxxxxxx.....
SEQ      hhhhhhhhhccccchhhhhhhccccceeeccccccccccccchhhhhhhhhhhccccccc
PRD      .....
MEM

```

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00007	78->86	TYR_PHOSPHO_SITE	PDOC00007
PS00007	77->86	TYR_PHOSPHO_SITE	PDOC00007
PS00008	97->103	MYRISTYL	PDOC00008

797

DKF2phtes3_2h15

group: testes derived

DKF2phtes3_2h15 encodes a novel 855 amino acid protein with very weak similarity to *S. pombe* cdc23.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to cdc23

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4619 bp

Poly A stretch at pos. 4598, polyadenylation signal at pos. 4589

```
1 GAAGCGCTCC CGGCATCGGC CAAGATTCTA CATTGCTCAT CTGGGCATCT
51 GAGCCTCCTT CGAAGTTTCC TGTCACAACT GTCCTCTTGA CAGCATGGAT
101 GAGGAGGAAG ACAATCTGTC TCTGCTGACC GCACTGCTGG AAGAAAATGA
151 GTCAGCCTTG GATTGTAATT CAGAAGAAAA TAACCTCTTG ACGCGGAAAA
201 ATGCCGAGCC CGACGCATTG GATGAGCTCT TTGATGCCGA CGGCGACGGT
251 GAATCTTATA CAGAAGAGGC TGATGATGGA GAAACAGGAG AGACAAGAGA
301 CGAAAAGGAA AATCTGGCCA CTCTCTTTGG AGATATGGAG GACTTAACAG
351 ATGAAGAGA AGTTCGCCA TCACAGTCAA CTGAAAATAG GGTCTCCCT
401 GCTCCTGCCC CCAGGCGAGA GAAAACGAAT GAAGAGTTGC AAGAGGAATT
451 AAGGAATTTG CAAGAGCAAA TGAAGGCTT ACAAGAGCAG CTAAGAGTAA
501 CAACAATTAA ACAGACAGCA AGCCCAGCCC GTCTGCAAAA ATCCCCTGAG
551 AAGTCTCCCC GGCCACCTCT TAAGGAGAGG AGAGTTCAGA GAATTCAGGA
601 GTCAACATGC TTTTCTGCGG AGCTTGATGT CCCTGCGCTA CCAAGAACCA
651 AGAGGGTGGC TCGAACACCA AAGCCTTCAC CTCAGATCC CAAAAGCTCA
701 TCTTCAAGGA TGACAAGTGC ACCCTCCCAA CCCCTACAGA CGATTTCTCG
751 GAACAAACCT AGTGGGATAA CTAGAGGTCA AATTGTGGGG ACCCCAGGAA
801 GTTCTGGGGA AACGACTCAA CCCATCTGTG TGGAAAGCTT CTCTGGTCTG
851 CGGCTCAGGC GGCCTCGAGT ATCCTCCACA GAAATGAACA AGAAAATGAC
901 CGGCCGAAAA CTGATCAGAC TGTCTCAGAT CAAGGAAAAAG ATGGCCAGAG
951 AGAAGCTGGA AGAATAGAT TGGGTGACAT TTGGGGTTAT ATTGAAGAAG
1001 GTTACGCCAC AGAGTGTGAA TAGTGGAAAA ACCCTTCAGCA TATGGAAACT
1051 GAATGATCTT CGTGACCTGA CACAATGTGT GTCCTTGTC TTATTGGAG
1101 AAGTTCACAA AGCGCTCTGG AAGACGGAGC AGGGGACTGT CGTAGGGATC
1151 CTCATGCGCA ACCCATGAA GCCCAAGGAT GGTTCAGAGG AGGTGTGTTT
1201 ATCTATCGAT CATCTCAGA AGGTCTTAAT TATGGGTGAA GCTCTTGACC
1251 TGGGAACCTG TAAAGCCAAG AAGAAGATG GAGAGCCGTG CACGCAGACT
1301 GTGAATTTGC GTGACTGTGA GTACTGTGAG TACCATGTCC AGGCTCAGTA
1351 CAAGAAGCTC AGTGCAAGC GTGCGGATCT GCAGTCCACC TTCTCTGGAG
1401 GACGAATTCC AAAGAAGTTT GCCCGCAGAG GCACCAGCCT CAAAGAACGG
1451 CTGTGCCAAG ATGGCTTTTA CTACGGAGGG GTTCTTCTG CCTCGTATGC
1501 AGCTTCAATT GCAGCAGCTG TGGCTCCTAA GAAGAAGATT CAAACCATC
1551 TGAGTAATCT GGTGTGTAAG GGCACAAACT TGATCATCCA GGAACACCGG
1601 CAAAACTCG GAATACCCCA GAAGAGCCTG TCTTGCTCTG AGGAGTTCAA
1651 GCAACTGATG GACCTGCCGA CGTGTGGAGC CAGGAACCTA AAACAACATT
1701 TAGCCAAAGC CTCAGCTTCA GSGATTATGG GGAGCCCAAA ACCAGCCATC
1751 AAGTCCATCT CGGCCTCAGC ACTCTTGAAG CAACAGAAGC AGCGGATGTT
1801 GGAGATGAGG AGAAGGAAAT CAGAAGAAAT ACAGAAGCGA TTTCTGCAGA
1851 GCTCAAGTGA AGTTGAGAGC CCAGCTGTGC CATCTTCATC AAGACAGCCC
1901 CCTGCTCAGC CTCCACGGAC AGGATCCGAG TTCCCCAGGC TGGAGGGAGC
1951 CCCGGCCACA ATGACGCCCA AGCTGGGGCG AGGTGTCTTG GAAGGAGATG
2001 ATGTTCTCTT TTATGATGAG TCACCACCAC CAAGACCAA ACTGAGTGCT
2051 TTAGCAGAAG CCAAAAAGTT AGCTGCTATC ACCAATTAA GGGCAAAAGG
2101 CCAGGTTCTT AAAAAACAA ACCCAAACAG CATTAAAGAG AAACAAAAGG
2151 ACCCTCAGGA CATCTGGAG GTGAAGGAC GTGTAGAAAA AAACACCATG
2201 TTTTCTCTC AAGCTGAGGA TGAATTGGAG CCTGCCAGGA AAAAAAGGAG
2251 AGAACAACCT GCCTATCTGG AATCTGAGGA ATTTAGAAA ATCTTAAAG
2301 CAAATCAAA ACACACAGGC ATCTGAAAG AGGCCGAGGC TGAGATGCAG
2351 GAGCGCTACT TTGAGCCACT GGTGAAAAAA GAACAAATGG AAGAAAAGAT
2401 GAGAAACATC AGAGAAGTGA AGTGCCGTGT CGTGACATGC AAGACGTGCG
2451 CCTATACCCA CTCAAGCTG CTGGAGACCT GCGTCAGTGA GCAGCATGAA
2501 TACCACGGC ATGATGGTGT GAAGAGGTTT TTCAATATGC CTTGTGGAAA
2551 CAGAAGCATC TCCTTGGACA GACTCCCGAA CAAGCACTGC AGTAACGTGT
2601 GCCTCTACAA ATGGGAACGG GACGGAATGC TAAAGGTATG CCATTTGCGT
2651 ACTAATTTTT GACTCCTTTT AGTGACCCAT GCTAATAATG TGAACCATC
```



```

2701 TCCTATTAAA ATATTTTCAT TTTTCTAGGA AAAGACTGGT CCAAAGATAG
2751 GAGGAGAAAC TCTGTTACCA AGAGGAGAAAG AACATGCTAA ATTTCTGAAC
2801 AGCCTTAAAT AACCCGAACT TCAGACATTT TCCCACAGAC TTCCTGGCCT
2851 CCTGTGACTC TGGAAAGCAA AGGATTGGCT GTGTATTGTC CATTGATTCC
2901 TGATTGACGC CGTCAAAAAC AAATGCTTGT TAAGCCCATA AGCTTTGCCT
2951 GCTTACTTTC TGCCATTGGG TTGGTTTGAT ACCACATTTA ACATTGACAT
3001 TTAAGTGGAA AACCAAGTTA TCATTGTCTT TCTAAGCTC AGTGTGGATG
3051 ATTGCAATTAC TTCATTCACT GAAGTTTTTG CCAAAAATT GGAAGGTAAA
3101 CAGAGAGCTA TGTTTCTGTA TCTTTGGTT ATAGAGTGTT CACTTCTTTA
3151 TCATAACAAA ATTCTAGTGT TTATACGAAC ACCCAGAGGC AAAAGAATTT
3201 GGCTTAAATC TCACTCCAGG TAAGTAGCTT AACTTCTGGG CTTCAGTTTT
3251 CTCATCTGTA AAATCAGGAA GATTGGACTA AGTGATCCTG AAATGTATTT
3301 TTTAGCACTG GATTCTACA AATAATAAAA CTTTCCCATC TAGATAATGA
3351 TGATCACATA GTCTTGATGT ACGGACATTA AAAGCCAGAT TTCTTCATTC
3401 AATTCTGTGA TCTCTGTTTT ACTCTTTGAA ATTGATCAAG CCAGTGAATC
3451 ACTTTGCATT TCAGTTTATA TATAGAGAGA GAAAGAAGGC TGTCTGCTCT
3501 TACATTATTG TGGAGCCCTG TGATAGAAAT ATGTAAAATC TCATATTATT
3551 TTTTTTTTAA TTTTTTTATT TTTTATGACA GGGTCTCACT ATGTCACCTT
3601 GGCTGGAGTG CAGTAGTGCG ATCGCGGCAC ACTGCAGCCT TGGCTTCCCT
3651 GGGCTCAAGC AGTCCTCCCA CCTCAGTCTC CCAATAGCT AGGACTACAG
3701 GCGTGCCTGA CCAAGCCAG CTAATTTTGT CATTTTGTGT AGAGATGGGG
3751 TTTTGGCATG TTGCTCAGGC TGGTCTCAAA CTCCTGAGCA CTAGCAATCC
3801 ACCCACCTCT GTTCCAAAA AAAAAAAAAA AATGAAAGGT CAACCCCTAT
3851 GCAAAATTACC ACAGCAAAGG TTTCATTGAG GAGATTCTTC CATCTGGGCA
3901 ACCTGGTTTT CCAAAATATCA TTTGACCTAA GTGAATGTGT ATACTAGCTA
3951 AAGATTGGGT AAATTGGTTG AATTATTGTA TTGAAGCTTG AGCTGTAGCT
4001 AAAAGTAATT TAGGTTTCCC CTAAGATGTT ATTATGTTAG GGACATAACA
4051 CTTTGGGAG GTTGTGTGGG GAGATGGTTG ATTTAGGTTT TCAAAAGCTA
4101 GAAATAAAAT TTACATGCCT TAGATTTCAT AAAATTCTGC TCTAATTGGG
4151 TGGAAGGTGC TGTATCTAAC TTGTGTTCCCT CCTAAGGTTA TGTCTTAATA
4201 ACTATTCTTT TAGGAGTATA CTTCTACTTT ATAGAAGGTT GCTTTTCTTT
4251 TTAATTTTTT CTAACAAGA AAAGAATAAA GTATTATTAT ATAAGAACCA
4301 GAAAGCACTT GAAACTGATG TTTTAAATGG CTCATTTAGG GTAGATTAT
4351 TTATCTCATT AACTTAAAC AGCTATGTGT ATGAAATAGG TCACAACAGA
4401 ACTTGAACAC CAGGTTGGTG TCTGAGCAAT CCCTTTCTTA TGGGAAAAAC
4451 AATGTTCTTG TTTGAACAGA GGGTATCATT GCAGTCAGTA TTCACGTGTA
4501 TATTGTTATA TAAGTTGTAT AATATGCTTG TAAAGGCTGA GGGTGAGCTG
4551 TATCTGGATG CCTTTTACA ATTTGATTTT AACTTTTAAA ATAAATTTAA
4601 AACATAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 95 bp to 2659 bp; peptide length: 855
 Category: similarity to known protein
 Classification: Cell division

```

1 MDEEEDNLSL LTALLEENES ALDCNSEENN FLTRENGEPD AFDELFDADG
51 DGESYTEEAD DGETGETRDE KENLATLFGD MEDLTDEEEV PASQSTENRV
101 LPAPAPRREK TNEELQEELR NLQEQMKALQ EQLKVTTIKQ TASPRLQKS
151 PEKSPRPPLK ERRVQRIQES TCFSAEILDVP ALPRTKRVAR TPKPSPPDPK
201 SSSSRMTSAP SQPLQTISRN KPSGITRGQI VGTPGSSGET TQPICVEAFS
251 GLRLRRPRVS STEMNKKMTG RKLIRLSQIK EKMAREKLEE IDWVTFGVIL
301 KKVTPQSVNS GKTFISIWKLN DLRDLTQCVS LFLFGEVHKA LWRTEQGTUV
351 GILNANPMKP KDGSSEEVCLS IDHPQKVLIM GEALDLGTCK AKKNGEPTC
401 QTVNLRDCEY CQYHVQAQYK KLSAKRADLQ STFGSGRIPK KPARRGTSLK
451 ERLCQDGFY GGVSSASYAA SIAAAVAPKK KIQTLSNLV VKGTNLIQE
501 TRQKLGIPOK SLSCSEEFKE LMDLPTCGAR NLKQHLAKAS ASGIMGSPKP
551 AIKSISASAL LKQKQRMLE MRRRKSEIQ KRFLQSSSEV ESPAVPSSSR
601 QPPAQPPRTG SEFPRLGAP ATMTPKLGRG VLEGDDVLFY DESPPRPKL
651 SALAEAKKLA AITKLRAKGQ VLTKTNPSNI KKKQKDPQDI LEVKERVEKN
701 TMFSSQAEDE LEPARKKRE QLAYLESEEF QKILKAKSKH TGILKEAEAE
751 MQERYFEPLV KKEQMEEKMR NIREVKCRV TCKTCAYTHF KLETCVSEQ
801 HEYHWDGVK RFFKPCGGR SISLDRLPNK HCSNCGLYKW ERDGMKVCH
851 LRTNF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h15, frame 2

TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347., N = 2, Score = 284, P = 7e-21

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7e-12

TREMBL:SCDNA52A_1 gene: "DNA52"; Saccharomyces cerevisiae DNA52 gene, complete cds., N = 2, Score = 201, P = 7.9e-12

TREMBLNEW:AC006234_6 gene: "F5H14.6"; Arabidopsis thaliana chromosome II BAC F5H14 genomic sequence, complete sequence., N = 2, Score = 211, P = 1.7e-15

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7.2e-12

>TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347.
Length = 593

HSPs:

Score = 284 (42.6 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 97/383 (25%), Positives = 186/383 (48%)

Query: 109 EKTNEELQEELRNLEQOMKALQEQLKVTTIKQTASPARLQKSPKSPRPPLKERRVQRIQ 168
E+ + +L+E + LQ Q+ +QE+ ++ + + AS + + PR P + + RV +
Sbjct: 8 EENDLDEE--KRLQRLNEIQEKKRLRSAQKEASSENAEVI--QVPRSPQQVRVLTVS 63

Query: 169 ESTCFSAE----LDVPALPRTKRVARTPKPSPDPKSSSSRMTSAPSQP-----LQTIS 218
+ + + L + + K V+ P P PK R+ A +Q L+T+
Sbjct: 64 SPSKLSKPKRLILGIDKGTGKDVSLGKGRGPLPKPHERLAEARNQERKRSCLKTKMK 123

Query: 219 RNKPSGITRGQIVGTPGSSGETTQPI-C--VEAFSGRLRRPRVSSTEMNKKMTGRKLIR 275
+N+ R + + G S E P+ C ++ +S + +S + + G ++
Sbjct: 124 KNRKQSFQRKRNILEDGKSEEEKFPMKDEIDPYSRQAIVIRYISDEVAKENIGGNQVYL 183

Query: 276 LSQIKEKMAREKLE--EID-WVTFGVILKKV-TPQSVNSGKTFSIWKLNDLRDLTQCVSL 331
+ Q+ + + K E E+D +V G++ T ++VN K + + L DL+ +C
Sbjct: 184 IHQLLLVRAPKFEAPEVDNYVMGIVASNSGTRETVNGNK-YCMLTLTDLKWQLEC--- 239

Query: 332 FLFGEVHKALWKEQGTVVGILNANPMKPKDGS--EEVCLSIDHPQKVLII-MGEALDLGTC 389
FLFG+ + WK + GTV+ +LN +KPK+ L +D VL+ +G + LG C
Sbjct: 240 FLFGKAFERYWKIQSGTVIALLNPEVLKPKNPDIGRFSCLKDSEYDVLLEIGRSKHLGYC 299

Query: 390 KAKKKNGEPTQTQTNLRDCEYCOYHVQAQYKKLSAKRADLQSTFSGGRIPKKFARRGTS 449
+++K+GE C ++ R + C+YHV ++ + R + S+ + P+ ARR
Sbjct: 300 SSRKSGELCKHWDKRGDOVCEYHVDLAVQRSMSTRTEFASSMATMHEPR--ARR---- 353

Query: 450 KERLCQDGF--YYGGVSSASYAASIAAAVAPKKKIQT 484
++R GF Y+ G ++ +A + +QT
Sbjct: 354 EKRRFGQGFQGYFAGEKYSAIPNAVAGLYDAEDAVQT 390

Score = 41 (6.2 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 12/43 (27%), Positives = 17/43 (39%)

Query: 453 LCQDGFYYGGVSSASYAASIAAAVAPKKKIQTTLNLNVKGTN 495
L +D S AS A++ K + SN + GTN
Sbjct: 465 LSKDSEIDSSTKKPSVLASFNASIMNPKSSLPSFSNSAILGTN 507

Score = 40 (6.0 bits), Expect = 8.9e-21, Sum P(2) = 8.9e-21
Identities = 13/26 (50%), Positives = 18/26 (69%)

Query: 536 LAKASASGIMGSPKPAIKSISASALL 561
LA +AS IM +PK ++ S S SA+L
Sbjct: 481 LASFNAS-IM-NPKSSLPSFSNSAIL 504

Pedant information for DKFZphtes3_2h15, frame 2

Report for DKFZphtes3_2h15.2

SEQ TCKTCAYTHFKLLETQVSEQHEYHWHHDGVKRFKPCGNRSISLDRLPNKHCSNCGLYKW

WO 01/12659

PCT/IB00/01496

SEG
PRD eeecceeeeeeccccceccccccccceeeccccccccccccccccccccceec
COILS

SEQ ERDGMLKVCHLRTNF
SEG
PRD ccccccccccccccc
COILS

(No Prosite data available for DKFZphtes3_2h15.2)

(No Pfam data available for DKFZphtes3_2h15.2)

DKF2phtes3_2i5

group: testes derived

DKF2phtes3 2i5 encodes a novel 151 amino acid protein with weak similarity to. C.elegans
cosmid F20D12.3

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific
genes.

similarity to C.elegans F20D12.3

many ATGs in front of the start of the ORF,
unspliced intron in 5' region?

Sequenced by EMBL

Locus: unknown

Insert length: 2142 bp

Poly A stretch at pos. 2121, polyadenylation signal at pos. 2102

```

1 GCAGTAAATA TGATATGAAA GAATTCTCTA ACTTGGGGGT GGCTTGTAAAC
51 CTGTAATAAA AATATTGCTA AAATACCTTC TCTCAGTTTG AAAAAGCATC
101 TGAGCAATCC TCAGTTATTG GTGAATTCCTT ACCAGTGTAT AATCCTCTC
151 TTTCCGTAT GGTCTTAGTG TGGTGTCTCT GGTGTAGTAT TTCAAGAGGA
201 ACCTGCAGCA AGATGAAAAG AGAGTGGGAC TTGGAGCTAA GAACGTTTTT
251 GGCTTTAAGT GCTACGTTAA CTCATTAAAT TCTTAGTGAT CTTGGGGAAG
301 TCCCTCACC AGTGTGAGCC TCAGTTTCTT TATCTAATAA GTAAGGATAA
351 TCTTACCCAC CTTATTGCGG GGGCCCGAGG ATTACATGAT TGGTGTAAAC
401 GTAGCACCTT GTACATTGTA AAGGACTAAT ACCAGTGGAC TTAACTCTTG
451 GCTGGGCTTT GGAATTCCTG GTGGGACTTT TTAATCATGT AGATTCTCAG
501 GCCCTGCCTT GGCTGTGGA ACCACAGACT CTATAGGTGG GCCCTTCCAG
551 AAGGCTTCAT GGGTGGTTCT CATGTGGAAC CTGTGTTGCA AGCCACTGCA
601 TGGTGTACT GCTATTAACA TTAATACTTA TATTTCTCTT ATTGTGTGGA
651 TATATCTGTG GTGTTTGCCC ATGTATACCTT CATTTTACAT TTCTTAAAGA
701 ATAGAATGGA ATGGTTTAA GCACGCTACA TTGTCCAGGT TATACCCACA
751 GAAGAGCTGT TGTGTAACAG AATCAGCATC ATACCTGAAT CATTGTGTACA
801 TTGCATATAA GACTATGTCT AAGTAGAAGA TGCTATGAAA TCATGTCTGC
851 TGTGGGGCCA GGCATAATTA TGAATGTTAC TTAAGAGCAT AGGTGAGGTG
901 AGAAAAGGGA ATGTGACTAG TGTTTAGTA TTTCTTGGT GTGGGATGAA
951 GTATAATCTT TTTTTTTTTT TCTCAACAAA GCAGTAAAC TAGAAAGAG
1001 GAGAACTCTT CCCTCAAGAA TGGCTGTACC TTCATATCTA GAGGCACATT
1051 AAAAAAAGA ACGTCTGTAC CTTAAAAATG GAGGTCATTT CATTGTGTTC
1101 ATTTTCAAGG TTGTTGTATG GCTCGGTGAG AACTTTCTGT TACCAGAAGA
1151 CACTCACATT CAGAATGCTC CATTTCAGT GTGTTTCACA TCTTTACGGA
1201 ATGGCGGCCA CCTGCATATA AAAATAAAAC TTAGTGGAGA GATCACTATA
1251 AATACTGATG ATATTGATTT GGCTGGTGAT ATCATCCAGT CAATGGCATC
1301 ATTTTGTGCT ATTGAAGACC TTCAAGTAGA AGCGGATTTT CCTGTCTATT
1351 TTGAGGAATT ACGAAAGGTG CTAGTTAAGG TGGATGAATA TCATTCACTG
1401 CATCAGAAGC TCAGTGCTGA TATGGCTGAT CATTCTAATT TGATCCGAAG
1451 TTTGCTGGTC GGAGCTGAGG ATGCTCGTCT GATGAGGGAC ATGAAAACAA
1501 TGAAGAGTCG TTATATGGAA CTCTATGACC TTAATAGAGA CTTGCTAAAT
1551 GGATATAAAA TTCGCTGTAA CAATCACACA GAGCTGTGG GAAACCTCAA
1601 AGCAGTAAAT CAAGCAATTC AAAGAGCAGG TCGTCTGCGG GTTGGAAAAC
1651 CAAAGAACCA GGTGATCACT GCTTGTGCGG ATGCAATTCG AAGCAATAAC
1701 ATCAACACAC TGTTCAAAAT CATGCGAGTG GGGACAGCTT CTTCTAGGT
1751 GAGGAAAATA CAGGTCATGA AGTTCCTGGC AAAGATTTTC TGTAAAAAAC
1801 CTATGCTGGT TTGCTTTGGA TCACACCTGT GTGAACCCCG GGTGCTAAGA
1851 ATGAAATAAA CCTTGGTGAG TTGTACAAAT TAAAGACAAA GAACTACATG
1901 TGAAGATAGA CTTGCTTCTT ATTTTAAAT CAGTAGTAGT ACTGTTGCTG
1951 AATAATACTA GGTTTTATG GAATAGGATG AATGCTTTTG AAGTATTAGG
2001 GCTTCAGAGT CCAATTTTGC TTATTTATGG TATATAAATA CATATTTT
2051 TCTTGAATTT GCAATTGAGT TTGTACTTTT CAAATAGATT ATCTACTTTT
2101 TCATTAAAT GTAAAGATGT TAAAAAATA AAAAAAATA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 1293 bp to 1745 bp; peptide length: 151
Category: similarity to unknown protein
Classification: no clue

```

1 MASFFAIEDL QVEADFPVYF EELRKVLVKV DEYHSVHQKL SADMDHNSNL
51 IRSLLVGAE D ARLMRDMKTM KSRYMELYDL NRDLNGYKI RCNNHTELLG
101 NLKAVNQAIQ RAGRLRVGKP KNQVITACRD AIRSNNINTL FKIMRVGTAS
151 S

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2i5, frame 3

TREMBL:CEF20D12_1 gene: "F20D12.3"; *Caenorhabditis elegans* cosmid F20D12., N = 1, Score = 173, P = 4.5e-12

```
>TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12.  
Length = 699
```

HSPs:

Score = 173 (26.0 bits), Expect = 4.5e-12, P = 4.5e-12
Identities = 33/130 (25%), Positives = 72/130 (55%)

```

Query:      20 FEELRKVLVKVDEYHSVHQKLSADMDHNSLIRSLLVGAEDARLMRDMTKMSRYMELYD 79
             F+E ++L ++D V ++L++Q ++ + ++ AED+ ++ + Y+ L
Sbjct:     569 FKADEILLEEIDPMTEVRLDLTAELQGRQAAVKETIIRAEDSIADNIPDARKFYIRLKA 628

Query:      80 LNRDLLNGYKIRCNNHTELLGNLKAVNQAIQIRAGRLRVGKPKNQVITACRDAIRSNNINT 139
             + ++R NN + +L+ +N+ I+ RLRVG+P Q++ +CR AI +N
Sbjct:     629 NDAARQAQLRWNNQERCVKSLRRLNKIIENC SRLRVGEGRQIVVSCRSIAIDDNKQI 688

Query:     140 LFKIMRVGT 149
             + KI++ G +
Sbjct:     689 ITKILQYGAS 698

```

Pedant information for DKFZphtes3_2i5, frame 3

Report for DKFZphtes3_2i5.3

```
[LENGTH]      151
[MW]           17304.07
[pI]           9.33
[HOMOL]        TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12. 2e-12
[KW]           Alpha Beta
```

[illegible]

(No Prosite data available for DKFZphtes3 2i5.3)

(No Pfam data available for DKFZphtes3 2i5.3)

DKFZphtes3_2119

group: testes derived

DKFZphtes3_2119 encodes a novel 166 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, no EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1079 bp

Poly A stretch at pos. 1053, polyadenylation signal at pos. 1038

```
1 CCACAGGACA CACTGTTCCC AGGGCACAGA CACCCTGGGC TTTGGTTGGG
51 TCTTGGCCTC CAGGTAGGGC CCTGTTGGGC AGCGGGCAGC AACTCCTGAG
101 ACACTACTGT GATTCTTGGT GGTGGCTGTG GTAAAAAACC TGCAGGGCTA
151 GAGTTTGGGG TGAGATTGAG CAGTAACTGT GGCCTCTCCT AGTGACAGTA
201 TGTCACCTCC ACTCCAGCA CGCATGCCCA CAGGCCACGG CCTCCACATC
251 ACAAAACCCC CACCAAGTTG CCCATCTATG GAGCAGCTCC CATACGGCAG
301 GGTGAGGCTC TTACCTCCAC CTCCAGGGCA CAGACAGGGG GAGCTCTGTC
351 TCACGTGAAG GCAATGAGGA GAGTTGAGGG CCCAGACCAG GCTAGGGGCC
401 ATCCCTTTC CCGAGCAGGC CTCAGGGAAG GACCAGCCCC ATTCCCATCT
451 GACCTAGCTC TTAGCCCAGG AGCCTGCATA GGGGAAGAAAG GACAGACAGG
501 GCCTCCTTAC TGGCTGACAC TCAGGAGGGG CTGGGGCAAG AGAGCAGAGG
551 GAGCGCAGGG CCAGGCAGGG GCTGCTGAGG ATCCATGGGA GCTCAGGGTG
601 CACAAAGGGG CTGCCCTTCC TGGGCTGAGG GCAGCATCCC TATGGGAGCT
651 GAGAAAGTCC AATCCTGAGA TGGGACAGTG CTGCCAGGG GTGTGTGGCT
701 GGGCCCTGAC AACAGTCTCC CCAAAGTGA CCACATCACC AGGCTCAGTT
751 CCAGGAAGGC TGAGAAAGTC CCAGTACACT GAGGATGCAC CTCAGTTACA
801 TAAAATAAAT GAAACTGGAG TACTAACGTA CAGTTTAAAG GTTATAGTTA
851 CTATTTTAT ATGATATACT AGTAATTTT GAATAGGGTA AACTTTAGGT
901 GTTTTGACAC CAAAAGAAAA CTACATGAGT TCATGCATGT GTTAAATTGC
951 TTTACTGTAG TAATCATTTA CATGTATATG TATATATGAA TATAATTATG
1001 GGCTCATTA AATTTAATAT TATAAATAGG TGACAAAGAA TAAAGTTAAC
1051 TGGAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 364 bp to 861 bp; peptide length: 166

Category: putative protein

Classification: no clue

```
1 MRRVEGPDQA RGHPLSRAGL REGPAPFSPD LGLSPGACIG KKGQTGPPYW
51 LTLRRGWGKR AEGAQQQAGA AEDPWELRVH KGAALPGLQA ASLWELRKS
101 PEMGQCCPGV CGWALTTVSP KVTTSPPGVP GRLRSAQYTE DAPQLHKINE
151 TGVLTYSCLKV IVTIFI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2119, frame 1

Report for DKFZphtes3_2119.1

[LENGTH] 166
[MW] 17691.35
[pI] 9.54
[KW] All_Beta
[KW] LOW_COMPLEXITY 7.23 %

SEQ MRRVEGPDQARGHPLSRAGLREGPAPFPFSDLGLSPGACIGKKGQTGPPYWLTLRRGWGKR
SEG
PRD ccc

SEQ AEGAQGQAGAAEDPWELRVHKGAAALPGLQAASLWELRKSNPEMGQCCPGVCGWALTTVSP
SEG xxx
PRD ccc

SEQ KVTTS PGSVPGRLRSAQYTEDAPQLHKINETGVLTYS LKVI VTI FI
SEG
PRD ccc

(No Prosite data available for DKFZphtes3_2119.1)

(No Pfam data available for DKFZphtes3_2119.1)

DKFZphtes3_2ml8

group: nucleic acid management

DKFZphtes3_2ml8 encodes a novel amino acid protein, with similarity to mouse Dhml.

The protein seems to play a role in nucleotide metabolism, RNA metabolism, but also in DNA repair and cell cycle. The yeast homologue is a DNA strand exchange protein required for sporulation and homologous recombination.

The novel protein can find application as multifunctional nuclease / exoribonuclease.

nearly identical to mouse Dhml

complete cDNA, complete cds, start at Bp 42, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3022 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2981

```
1 CTCGTCAGCC GGTGCGCCGC CGCCTCCAGC CGTGTGCCGC TATGGGAGTC
51 CCGGCGTTCT TCCGCTGGCT CAGCCGCAAG TACCCGTCCA TCATAGTCAA
101 CTGCGTGGAA GAGAAGCCAA AAGAATGCAA TGGTGTAAGG ATTCCAGTTG
151 ATGCCAGTAA ACCTAATCCA AATGATGTGG AGTTTGATAA TCTGTATTGG
201 GATATGAATG GAATCATCCA TCCCTGTACT CATCCTGAAG ACAAAACCAGC
251 ACCAAAAAAT GAAGATGAAA TGATGCTTGC AATTTTGTAG TACATTGACA
301 GACTTTTCAG TATTGTAAGA CCAAGAAGAC TTCTCTACAT GGCAATAGAT
351 GGAGTGGCAC CACGTGCTAA AATGAACCAG CAGCGTTCAA GGAGGTTTCT
401 GGCATCAAAA GAAGGAATGG AAGCAGCAGT CGAGAAGCAG CGAGTCAGGG
451 AAGAAATATT GGCAAAAGGT GGCTTCTTTC CTCCAGAAGA AATAAAAGAA
501 AGATTGTACA GCAACTGTAT TACACCAGGA ACTGAATTCA TGGACAATCT
551 TGCTAAATGC CTTGCTGATT ACATAGCTGA TCGTTTAAAT AATGACCCCTG
601 GGTGGAAAAA TTTGACAGTT ATTTTATCTG ATGCTAGTGC TCCTGGTGAA
651 GGAGAACATA AAATCATGGA TTACATTAGA AGGCAAGAGC CCCAGCCTAA
701 CCATGACCCA AATACTCATC ATGTTTATG TGGAGCAGAT GCTGATCTCA
751 TTATGCTTGG CCTTGCCACA CATGAACCGA ACTTTACCAT TATTAGAGAA
801 GAATTCAAAC CAAACAAGCC CAAACCATTG GGTCTTTGTA ATCAGTTTGG
851 ACATGAGGTC AAAGATTGTG AAGGTTTGCC AAGAGAAAAG AAGGGAAAAGC
901 ATGATGAAGT TGCCGATAGT CTTCTTGTG CAGAAGGAGA GTTTATCTTC
951 CTTGCGCTTA ATGTTCTTCG TGAGTATTGG GAAAGAGAAC TCACAATGGC
1001 CAGCCTACCA TTCACATTG ATGTTGAGAG GAGCATTGAT GACTGGGTTT
1051 TCATGTGCTT CTTGTGGGGA AATGACTTCC TCCCTCATTT GCCATCGTTA
1101 GAGATTAGGG AAAATGCAAT TGACCGTTTG GTTAACATAT AAAAAAATGT
1151 GGTACACAAA ACTGGGGGTT ACCTTACAGA AAGTGGTTAT GTCAATCTGC
1201 AAAGAGTACA GATGATCATG TTAGCAGTTG GTGAAGTTGA GGATAGCATT
1251 TTTAAAGAGA GAAAGGATGA TGAGGACAGT TTTAGAAGAC GACAGAAAGA
1301 AAAAAGAAAG AGAATGAAGA GAGATCAACC AGCTTTCAC TCTAGTGGAA
1351 TATTAATCC TCATGCCTTG GGTCAAGAA ATTCAACAGG TTCTCAAGTA
1401 GCCAGTAATC CGAGACAAGC AGCCTATGAA ATGAGGATGC AGAATAACTC
1451 TAGTCCTTCG ATATCTCCTA ATACGAGTTT CACATCTGAT GGCTCCCCGT
1501 CTCCATTAGG AGGAATTAA GAAAAGCAG AAGACAGTGA CAGTGAACCT
1551 GAGCCAGAGG ATAATGTCAG GTTATGGGAA GCTGGCTGGA AGCAGCGGTA
1601 CTACAAGAAC AAAATTGATG TGGATGCAGC TGATGAGAAA TTCCGTCGGA
1651 AAGTTGTGCA GTCGTACGTT GAAGGACTTT GCTGGGTTCT TAGATATTAT
1701 TACCAGGGCT GTGCTCTCCT GAAGTGGTAT TATCCATTTC ATTATGCACC
1751 ATTTGCTTCA GACTTTGAAG GCATTGCAGA CATGCCATCT GATTTTGAGA
1801 AGGGTACGAA ACCGTTTAAA CCACTAGAAC AACTTATGGG GGTATTTCCA
1851 GCTGCAAGTG GTAATTTTCT ACCTCCATCA TGGCGGAAGC TCATGAGTGA
1901 TCCTGATTCT AGTATAATTG ACTTCTATCC TGAAGATTTT GCTATTGATT
1951 TGAATGGGAA GAAATATGCA TGGCAAGGTG TTGCTCTCTT GCCATTCTGT
2001 GATGAGCGAA GGCTACGAGC TGCCCTAGAA GAGGTATACC CAGACCTCAC
2051 TCCAGAAGAG ACCAGAAGAA ACAGCCTTGG AGGTGATGTC TTATTGTGG
2101 GGAAACATCA CCCACTCCAT GACTTCATTT TAGAGCTGTA CCAGACAGGT
2151 TCCACAGAGC CAGTGGAGGT ACCCCCTGAA CTATGTCATG GGATTCAAGG
2201 AAAGTTTCTT TTGGATGAAG AAGCCATTCT TCCAGATCAA ATAGTATGTT
2251 CTCCTGTTCC TATGTTAAGG GATCTGACAC AGAACACTGT AGTCAGTATT
2301 AATTTTAAAG ACCCACAGTT TGCTGAAGAT TACATTTTAA AAGCTGTAAT
2351 GCTTCCAGGA GCAAGAAAGC CAGCAGCAGT ACTGAAACCT AGTGACTGGG
2401 AAAAATCCAG CAATGGACGG CAGTGGAAAG CTCAGCTTGG CTTTAACCGT
2451 GACCGGAGGC CTGTGCACCT GGATCAGGCA GCCTTCAGGA CTTTGGGCCA
2501 TGTGATGCCA AGAGGCTCAG GAACTGGCAT TTACAGCAAT GCTGCACCAC
2551 CACCTGTGAC TTACCAGGGA AACTTATACA GGCCGCTTTT GAGAGGACAA
2601 GCCCAGATT CAAAACCTAT GTCAAAATATG AGGCCCCAGG ATTCTTGGCG
2651 AGGTCCCTCT CCCCTTTTCC AGCAGCAAGG GTTTGACAGA GCGGTGGGG
```

```

2701 CTGAACCTCT GCTCCCATGG AACCGGATGC TGCAAACCCA GAATGCAGCC
2751 TTCCAGCCAA ACCAGTACCA GATGCTAGCT GGGCCTGGTG GGTATCCACC
2801 CAGACGAGAT GATCGTGGAG GGAGACAGGG ATATCCAGA GAAGGAAGGA
2851 AATACCCTTT GCCACCACCC TCAGGAAGAT ACAATTGGAA TTAAGCTTTT
2901 GTAAAGCTTT CCCAAATCCT TTCATCATT TACAGTTTIA TGCTATTTGT
2951 GGAAAGATTT CCTTCTCAAG TAGTAGTTT TAATAAACT ACAGTACTTT
3001 GTGTAACAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

95192042:
 Characterization of cDNA encoding mouse homolog of fission yeast dhpl+
 gene: structural
 and functional conservation.

97361754:
 Cloning and characterization of mouse Dhml cDNA, a functional homolog
 of budding yeast
 SEP1.

Peptide information for frame 3

ORF from 42 bp to 2891 bp; peptide length: 950
 Category: strong similarity to known protein

```

1  MGVPAFFRWL SRKYPSTIIVN CVEEKPKECN GVKIPVDASK PNPNDVEFDN
51  LYLDMNGIIH PCTHPEDKPA PKNEDEMMVA IFEYIDRLFS IVRPRRLLYM
101  AIDGVAPRAK MNQQRSSRRF ASKEGMEAAV EKQVRREEIL AKGGFLPPPEE
151  IKERFDSNCI TPGTEFMDNL AKCLRYIAD RLNDPQWKN LTVILSDASA
201  PGESEHKIMD YIRQRQAQPN HDPNTHHCLC GADADLIMLG LATHEPNFTI
251  IREEFKPNKP KPCGLCNQFG HEVKDCEGLP REKKGKHDEL ADSLPCEAGE
301  FIFLRLNVLR EYLERELTMA SLPTTFDVER SIDDWVFMCF FVGNDFLPHL
351  PSLEIRENAI DRLVNIYKNV VHKTGGYLTE SGYVNLQVRVQ MIMLAVGEVE
401  DSIFKRRKDD EDSFRRQKE KRKRMRDQAP AFTPSGILTP HALGSRNSPG
451  SQVASNPQRA AYEMRMQNN SPSISPTSF TSDGSPSLG GIKRKAEDSD
501  SEPEPEDNVR LWEAGWKQRY YKNKFQVDAA DEKFRKVVQ SYVEGLCWVL
551  RYYYQGCASW KWIYYPFHYAP FASDFEGIAD MPSDFEKGTK PFKPLEQLMG
601  VFPAASGNFL PPSWRKLMSD PDSSIIDFYP EDFAIDLNGK KYAWQGVALL
651  PFVDERLRA ALEEYVPLDT PEETRRNSLG GDVLFVGGKH PLHDFILELY
701  QTGSTPEVEV PPELCHGIQG KFSLEDEAIL PDQIVCSVPV MLRDLTQNTV
751  VSINFDPQF AEDYIFKAVM LFGARKPAV LKPSDWEKSS NGRQWKPLQLG
801  FNRDRRPVHL DQAAFRTLGH VMFRGSGTGI YSNAAPPVPT YQGNLYRPLL
851  RGOAQIPKLM SNMRPQDSWR GPPPLFQQQR FDRGVGAEP LPPNRMLQTO
901  NAAFQPNQYQ MLAGPGGYPP RRDRGGRQG YPREGRKYPL PPPSGRYNWN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2ml8, frame 3

PIR:I49635 mouse Dhml protein - mouse, N = 1, Score = 4765, P = 0

PIR:S43891 dhpl protein - fission yeast (Schizosaccharomyces pombe), N
 = 3, Score = 1172, P = 2e-197

PIR:S20126 exoribonuclease RAT1 (EC 3.1.11.-) - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 1146, P = 3.8e-175

PIR:S72531 exonuclease II - fission yeast (Schizosaccharomyces pombe),
 N = 4, Score = 622, P = 4.2e-125

>PIR:I49635 mouse Dhml protein - mouse
 Length = 947

HSPs:

Score = 4765 (714.9 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 884/930 (95%), Positives = 895/930 (96%)

```

Query:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKCEKNGVKIPVDASKPNPNDFDNLVLDMNGIIH  60
Sbjct:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKCEKNGVKIPVDASKPNPNDFDNLVLDMNGIIH  60

Query:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR  120
Sbjct:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR  120

Query:    121  ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD  180
Sbjct:    121  A K GMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD  180

Query:    181  RLNNDPGWKNTLVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG  240
Sbjct:    181  RLNNDPGWKNTLVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG  240

Query:    241  LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE  300
Sbjct:    241  LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE  300

Query:    301  FIFLRNLVREYLERELTMASLPFTFDVERSIDDWVFMCFFVGNDFLPHLSLEIRENAI  360
Sbjct:    301  FIFLRNLVREYLERELTMASLPFTFDVERSIDDWVFMCFFVGNDFLPHLSLEIRENAI  360

Query:    361  DRLVNIYKNVVKHTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE  420
Sbjct:    361  DRLVNIYKNVVKHTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE  420

Query:    421  KRRKMRDQPAFTPSGILTPhALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPTSF  480
Sbjct:    421  KRRKMRDQPAFTPSGILTPhALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPTSF  480

Query:    481  TSDGSPSLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRRKVVQ  540
Sbjct:    481  TSDGSPSLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRRKVVQ  540

Query:    541  SYVEGLCWVLRYYYQGCASWKWYPFHYAPFASDFEGIADMSPDFEKGTKPFKPLEQLMG  600
Sbjct:    541  SYVEGLCWVLRYYYQGCASWKWYPFHYAPFASDFEGIADMSPDFEKGTKPFKPLEQLMG  600

Query:    601  VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA  660
Sbjct:    601  VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA  660

Query:    661  ALEEVYPDLTPEETRRNSLGGDVLVFGKHHPLHDFILELYQTGSTPEVPELCHGIQG  720
Sbjct:    661  ALEEVYPDLTPEETRRNSLGGDVLVFGKHHPLHDFILELYQTGSTPEVPELCHGIQG  720

Query:    721  KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAV  780
Sbjct:    721  KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAV  780

Query:    781  LKPSDWEKSSNGRQWKPLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPPVT  840
Sbjct:    781  LKPSDWEKSSNGRQWKPLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPPVT  840

Query:    841  YQGNLYRPLLRGQAQIPKLSNMNRPDQSWRGPPPLFQQQRFDRGVGAEPPLPWNRMLQTQ  900
Sbjct:    841  YQGNLYRPLLRGQAQIPKLSNMNRPDQSWRGPPPLFQQQRFDRGVGAEPPLPWNRMLQTQ  900

Query:    901  NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ  929
Sbjct:    901  NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ  929

```

Pedant information for DKFZphtes3_2m18, frame 3

Report for DKFZphtes3_2m18.3

```

[LENGTH]      950
[MW]           108582.68
[pI]           7.26
[HOMOL]        PIR:I49635 mouse Dhml protein - mouse 0.0
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YOR048c] 1e-123

```

```

[FUNCAT]      30.10 nuclear organization      [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]      01.03.16 polynucleotide degradation [S. cerevisiae, YGL173c] 3e-79
[FUNCAT]      30.03 organization of cytoplasm [S. cerevisiae, YGL173c] 3e-79
[FUNCAT]      03.22 cell cycle control and mitosis [S. cerevisiae, YGL173c] 3e-79
[PIRKW]       nucleus 1e-126
[PIRKW]       hydrolase 1e-122
[PIRKW]       exoribonuclease 1e-122
[PROSITE]     MYRISTYL 7
[PROSITE]     AMIDATION 2
[PROSITE]     CAMP_PHOSPHO_SITE 1
[PROSITE]     CK2_PHOSPHO_SITE 12
[PROSITE]     TYR_PHOSPHO_SITE 1
[PROSITE]     GLYCOSAMINOGLYCAN 1
[PROSITE]     PKC_PHOSPHO_SITE 8
[PROSITE]     ASN_GLYCOSYLATION 4
[KW]          TRANSMEMBRANE 1
[KW]          LOW_COMPLEXITY 6.21 %

```

```

SEQ      MGVPAFFRWLSRKYPISIVNCVEEKPKECNGVKIPVDASKPNPNDFDNLVLDMNGIIH
SEG
PRD      cccchhhhhhhhhccceeeeeecccccccccccccccccccccccccccccccccccccccc
MEM

```

```

SEQ      PCTHPEDKPAKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSRFR
SEG
PRD      cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      RLNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG
SEG
PRD      hccccccccceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
SEG
PRD      cccccccccccccccccccccccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      FIFLRLNVLEIRELTMASLPFTFDVERSIDDWVFMCFVGNDFLPHLPSLEIRENAI
SEG
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

```

SEQ      DRLVNIYKNVVKHTGGYLTESGYVNLQVRQIMLAVGEVEDSIFKKRKDDSDSFRRRQKE
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      KRKRMRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSSPSISPTSF
SEG
PRD      xxxxxxxx.....xxxxxxxxxxxxxx.....xxxxxxxxxxxxxx
MEM      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      TSDGSPSPGLGKIRKAEDSDSEPEPEDNVRLWEAGWKQRYYNKFVDVDADEKFRKRVVQ
SEG
PRD      xx.....xxxxxxxxxxxxxx.....
MEM      cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      SYVEGLCWVLRYYYQGCASWKWYYPFHYAPFASDFEGIADMPDSEKGTGKPKPLEQLMG
SEG
PRD      hhhhhhheeeeeeccccccccccccccccccccccccccccccccccccccccccccchhhhhh
MEM

```

```

SEQ      VFPAASGNFLPPSWRKLMSPDSSIIDFYPEDFAIDLNGKKYANQGVALLPFVDERRLRA
SEG
PRD      hccccccccccccccccccccccccceeeccccceeeccccceeeeeeccccchhhhhh
MEM

```

```

SEQ      ALEEVPDLTPEETRRNSLGGDVLFGVGHHPHLDHFILELYQTGSTPEVVPPELCHGIQG
SEG
PRD      hhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM

```

```

SEQ      KFSLEEAAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV
SEG

```

```

PRD      cccccceeeccccceccccccccccccccccccccccccchhhheccccccccce
MEM      .....

SEQ      LKPSDWEKSSNGRQWKPOLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPVVT
SEG      .....
PRD      eccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccc
MEM      .....

SEQ      YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEP LLPWNRMLQTQ
SEG      .....
PRD      cccccchhhhhccccchhhhhccccccccccccccccccccchhhhhccccccccccccchhhhhh
MEM      .....

SEQ      NAAFQPNQYQMLAGPGGYPPRRDRGGRQGYPREGRKYPLPPPSGRYNWN
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      hcccccccccecccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphtes3_2m18.3

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	247->251	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	477->481	ASN_GLYCOSYLATION	PDOC00001
PS00002	826->830	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	675->679	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00005	559->562	PKC_PHOSPHO_SITE	PDOC00005
PS00005	613->616	PKC_PHOSPHO_SITE	PDOC00005
PS00005	674->677	PKC_PHOSPHO_SITE	PDOC00005
PS00005	868->871	PKC_PHOSPHO_SITE	PDOC00005
PS00005	944->947	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	501->505	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	619->623	CK2_PHOSPHO_SITE	PDOC00006
PS00006	624->628	CK2_PHOSPHO_SITE	PDOC00006
PS00006	670->674	CK2_PHOSPHO_SITE	PDOC00006
PS00006	723->727	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->667	TYR_PHOSPHO_SITE	PDOC00007
PS00008	125->131	MYRISTYL	PDOC00008
PS00008	375->381	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00008	825->831	MYRISTYL	PDOC00008
PS00008	829->835	MYRISTYL	PDOC00008
PS00008	926->932	MYRISTYL	PDOC00008
PS00009	638->642	AMIDATION	PDOC00009
PS00009	934->938	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2m18.3)

DKFZphtes3_2m20

group: testes derived

DKFZphtes3_2m20 encodes a novel 183 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

group: unknown

DKFZphtes3_2m20 encodes a novel

amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

EST hits are only from testis or uterus librarys
remaining intron in3' UTR see EST-BLAST

Sequenced by EMBL

Locus: unknown

Insert length: 1341 bp

Poly A stretch at pos. 1320, polyadenylation signal at pos. 1300

```
1  GCAATCCAGG AGCTGAATGG TAACTCTTCC ACAAGCGAAA ACTGTTTCGTG
51  AATACAAGCA AAAGGCCCCC CAAGAGGACC CCTGATATGA TCCAGCAGCC
101 TCGGGCCCCG CTGGTGTGGG AGAAGGCTTC TGGTGAAGGA TTTGGCAAAA
151 CCGCCGCTAT TATACAGCTC GCTCCTAAAG CTCCTGTTGA CCTGTGTGAG
201 ACAGAGAAAC TGAGGGCAGC CTTCTTTGCA GTCCCGTTGG AAATGAGAGG
251 GTCCTTCCTG GTGCTGCTCC TGAGGGAATG CTTCCGAGAC CTGAGCTGGC
301 TGGCACTCAT CCATAGCGTC CGTGGGGAGG CGGGGCTGCT GGTGACGAGT
351 ATCTGCCCCG AGACCCCGTT TTTCTGGGCC ATGCACATCA CTGAGGCTCT
401 GCACAGAAC ATGCAGGCTC TGTTTAGCAC CCTGGCTCAG GCGGAGGAGC
451 AGCAGCCCTA CCTGGAGGCT CCACCGTTAT CGCGGGGACT CGCTGTCTGG
501 CAGAGTACCA CCTGGGGGAT TATGGACACG CCTGGAACAG GTGTGGGTG
551 CTGGACAGGG TGGACACCTG GGCTGTGGTC ATGTTCAATTG ATTTTGGACA
601 GTTGGCCACC ATCCCTGTGC AGTCTCTGCG CCAGCTAGAC AGCGACGACT
651 TCTGGACCAT CCCACCCCTG ACTCAGCCAT TCATGCTGGA GAAAGACATT
701 TTGAGTTTCG ATGAGGTTGT CCATCGAATC CTCAAAGGGA AAATCACTGG
751 TGCTTTGAAC TCGGCGGTAA CTGCTCCTGC ATCTAACTTG GCTGTTGTCC
801 CTCCTACTCT GCCCTTGGGG TGTCTGCAGC AGGCTGCTGC CTAGGCCTGG
851 ACACATTGCA CATCTAAAG TTTGAAGAGT CTAATAACG GGGCTTCCCT
901 CAGCATGTTT CCTCTCTGT TTGCCAGGA TCCAGAGCCA CTGCCCCTGT
951 CTTCTCGTAC CCCTTTCCT CTTGAGGCTT GGGAGGTGAA AAAGGCCAGA
1001 CTGTGCCAG GATTGATTCA ATTTTGCTTT TACTCCCAGC TTCCCTCTCA
1051 AAAGAGAGTG AAGTCTCATT TGTCTGTGTG CTTCAAGTTC CCAACTTGGC
1101 ATGAACATTG GAACCAACA TAGGAAACTA CCATTAGGTT GAAAGCCTGA
1151 GGCAGCTGGG ATGGTCTTTC TTGTGTCTCT TCTTTGCACC CCAGAGCATG
1201 ATATAAGTGG TCCTAACAGA TTCTGGATAA TGGAGAAGCC CTCTGCTGGT
1251 TTTCTGGCA TTCCATGTAG AATAGGTAGA GAATATTAA CCAATGAGCA
1301 AATAAATGTT GGCATGTTT ATGAAAAAAA AAAAAAAAAA A
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 479 bp to 841 bp; peptide length: 121
Category: questionable ORF
Classification: no clue

1 MRGTRCLAEY HLG DYGHAWN RCWVLD RVD T WAVVMFIDFG QLATIPVQSL
51 RQLDSDDFWT IPPLTQPFML EKDILSSYEV VHRILKGKIT GALNSAVTAP
101 ASNLAVVPPL LPLGCLQQA A

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 87 bp to 635 bp; peptide length: 183
Category: putative protein
Classification: no clue

1 MIQQPRAPLV LEKASGEGFG KTAATIQ LAP KAPVDLCETE KLRAAFFAVP
51 LEMRGSFLVL LLRECFRDL S WLALIH SVRG EAGLLVTSIV PKTPFFWAMH
101 ITEALHQNMQ ALFSTLAQAE EQQPYLEAPP LCAGLAVWQS TTWGIMDTPG
151 TGVGCWTGWT PGLWSCSLIL DSWPPSLCSL CAS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2m20, frame 2

Report for DKFZphtes3_2m20.2

[LENGTH] 121
[MW] 13436.69
[pI] 5.81
[KW] Alpha_Beta

SEQ MRGTRCLAEYHLGDYGHAWNRCWVLD RVD T WAVVMFIDFGQLATIPVQSLRQLDSDDFWT
PRD ccchhhhhcc

SEQ IPPLTQPFMLEKDILSSYEVVHRILKGKITGALNSAVTAPASNLAVVPPL LPLGCLQQA
PRD cccccchhhhhcchhhhhhhhhccccchhhhhcccccccccccccccccccccccccc

SEQ A
PRD C

(No Prosite data available for DKFZphtes3_2m20.2)

(No Pfam data available for DKFZphtes3_2m20.2)

Pedant information for DKFZphtes3_2m20, frame 3

Report for DKFZphtes3_2m20.3

[LENGTH] 183
[MW] 19971.49
[pI] 5.31
[KW] Alpha_Beta

```

SEQ  MIQQPRAPLVLEKASGEGFGKTAIIQLAPKAPVDLCETEKLRAAFFAVPLEMRGSFLVL
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  LLRECFRDLSWLALIHVSVRGEAGLLVTSIVPKTPFFWAMHITEALHQNMQALFSTLAQAE
PRD  hhhhhhcchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc
SEQ  EQQPYLEAPPLCAGLAVWQSTTWGIMDTPGTGVGCWTGWTPGLWSCSLILDSWPPSLCSL
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  CAS
PRD  ccc

```

(No Prosite data available for DKFZphtes3_2m20.3)

(No Pfam data available for DKFZphtes3_2m20.3)

DKFZphtes3_2n9

group: testes derived

DKFZphtes3_2n9 encodes a novel 184 amino acid protein with very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

on genomic level encoded by HS1186N24, no splice pattern but EST matches

Sequenced by EMBL

Locus: unknown

Insert length: 1000 bp

Poly A stretch at pos. 988, polyadenylation signal at pos. 970

```
1 CAACCTTTTAA AAGATGTGAA TTGGACAGCC AGACTTGCTT ATTTGTCTGA
51 TATCTTCAGT ATTTTAAAT GATCTTAATG CTTCTATGCA AGGGAAGAAT
101 GCAACTTATT TTTCATGGC AGATAAAGTT GAAGGACAAA AACAGAAGTT
151 AGAAGCTTGG AAAACAGAA TTTCTACAGA TTGTTATGAC ATGTTTCATA
201 ATTTAACAAC AATTATCAAT GAAGTAGGTA ATGATCTTGA TATTGCACAT
251 CTGCGAAAAG TTATCAGTGA ACATCTTACA AATTGTTAG AATGTTTTGA
301 ATTTTATTTT CCATCAAAAG AAGATCCACG CATAGGAAAT TTGTGGATCC
351 AAAATCCATT TCTTTCATCA AAAGATAACT TAAATTAAAC TGTAACTCTA
401 CAGGATAAGT TGTGAAGCT GGCTACCGAC GAAGGATTGA AAATCAGTTT
451 TGAAAATACA GCATCACITC CTTCAATTTG GATAAAAGCT AAAAATGACT
501 ATCTGAGCTT TGCTGAGATT GCTTTAAAT TGCTGCTTCT TTCCCTCA
551 ACATACCTCT GTGAGACCGG ATTCTCTACT TTAAGTTTA TTAACAAA
601 ACATAGAAAC AGTTTAAATA TACATTATCC CCTGAGGTAG CATTGTCAATC
651 AATCCAACCT AGATTAGACA AATTAACAAG CAAGAAGCAA GCTCACTTAT
701 CACATTAAAA GCTTTAAATA TTGATATGTA AGGTATTGGT TCAAAGTATG
751 CATATAAGCA TTGAGTGTGA GGAATTTGCT ATTTCACTTT AACTTTCTG
801 TCTAGTTACA GTTATGGAAG TATGAGAAGT TATGAGTGAA ACAGCAATTT
851 TCTATATAAA TTGCCTATAT GTATATTTTC AATTAAGAAAT GTGTACAGTT
901 TTTATAATTC TATTTTCTCT CATATTGTGC GTATTTATTA AAATATAATT
951 TTAATCTCTG TGATTCTAAT ATTAAACAT TTGATCTTAA AAAAAAAAAA
```

BLAST Results

Entry HS1186N24 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 1186N24
Score = 4921, P = 5.8e-215, identities = 989/992

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 86 bp to 637 bp; peptide length: 184
Category: similarity to unknown protein
Classification: no clue

```
1 MQGKNATYFS MADKVEGQK KLEAWKNRIS TDCYDMFHNL TTIINEVGND
51 LDIAHLRKVI SEHLTNLLEC FEFYFPSKED PRIGNLWIQN PFLSSKDNLN
101 LTVTLQDKLL KLATDEGLKI SFENTASLPS FWIKAKNDYP ELAEIALKLL
151 LLFPSTYLCE TGFSTLSVIK TKHRNSLNIH YPLR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2n9, frame 2

TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence., N = 1, Score = 94, P = 0.042

>TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence.
Length = 533

HSPs:

Score = 94 (14.1 bits), Expect = 4.3e-02, P = 4.2e-02
Identities = 39/177 (22%), Positives = 75/177 (42%)

```
Query:   1 MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLD-IAHLRKV 59
      +QG +   M D +   KL W+ ++ +   F L   + L+ I + ++
Sbjct:  354 LQGSQIVTQMYDLIRAF LAKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYPKIAEL 413

Query:   60 ISEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLLSSKDNLNLTVTLQDKLLKLATDEGLK 119
      +E   L + F+ Y + + +   +PF + D+++ LQ +++ L + LK
Sbjct:  414 KTEFQKRLSD-FKLY---ESELTL---FSSPFSTKIDSVH--EELQMEVIDLQCNTVLK 463

Query:   120 ISFENTASLPFWIKAKNDYPXXXXXXXFPSTYLCETGFSTLSVIKTKHRNSL 177
      ++   +P F+   YP   F STY+CE FS + + KTK+ + L
Sbjct:  464 TRYDKVG-IPFQYKYLWGSYPKYKHHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQL 520
```

Pedant information for DKFZphtes3_2n9, frame 2

Report for DKFZphtes3_2n9.2

```
[LENGTH]      184
[MW]           21203.53
[pI]           6.52
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      6.52 %
```

```
SEQ  MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLDIAHLRKVI
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhcchhhhhccceccccccccchhhhhhhhh

SEQ  SEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLLSSKDNLNLTVTLQDKLLKLATDEGLKI
SEG  .....
PRD  hhhhhhhhhhhhhccccccccccccceccccccccccccceehhhhhhhhhhhcccee

SEQ  SFENTASLPFWIKAKNDYPELAELKLLLPSTYLCETGFSTLSVIKTKHRNSLNIH
SEG  .....
PRD  eccccccccceccccchhhhhhhhhhhhhccccccccccccceccccccccceec

SEQ  YPLR
SEG  ....
PRD  cccc
```

(No Prosite data available for DKFZphtes3_2n9.2)

(No Pfam data available for DKFZphtes3_2n9.2)

DKFZphtes3_30f4

group: testes derived

DKFZphtes3_30f4 encodes a novel 192 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by LMU

Locus: /map="717.2-8 cR from top of Chr8 linkage group"

Insert length: 1388 bp

Poly A stretch at pos. 1330, polyadenylation signal at pos. 1310

```
1  CACTGAGCCC TCCTCAGATG GTTAGTGGCT TCCAACAGCC ATCAGGAGTG
51  TTTCTTGAAT GCCCCAGGTG TGGAGGACTT GGTCTGTGAC CACCTAGAAC
101 CCCAGAGCTG AACAGGAAGC CGTCCCTGCA GCAACAAGAG GGCTGGAAGG
151 GGGAGCTGCA GGCCACCCCTC GGCTCTCCCA CTGCTGGGGC GGTGATGTTT
201 GGGTGACATG TTTGAAAAAT ACTCTTAAAG ATACCAACTG TTCCCTTATA
251 TGGCTAATGG TTTGTGCAGC CACCAGCGAT GCGGGCCCTT ATTAGAGACC
301 AGGTTTGTGA AAACACCAAA TATTGCTGTC CACACTAGAC ATTAACCGGC
351 TTCAGAAAAAG ATGGACACCT TTTCCACACG TGTTTCGCTT CTTAACTTTG
401 GTCCAGCTTT AGCCACCACA CAGCGTGTGA GGGACTGCTG CTGCGGAGTC
451 AGCCTCGTTT GTCCCTCCGC CTCCCACCAG CATGCGCCGC TTCTGAGAGA
501 CACCAGCTCC CTGCTCCCAA GCCTGGTGCC ACAGGCCTGT CGTGAGGGAC
551 CCCTGCTTCC GAGAGCTCCT GGGGGGGTTC TGCCCTTCAC CACCTGGGAG
601 AGGTGTCTAG TCAGTTCCGA GTTGAACAAG GCCCGTGAC ACAGCATGTT
651 GGGGGGCCAG CCCAAAGTTC TTGTACCTC CTCATGCAA GCCAGCCATC
701 ACCCTCCGGC CAGAGCTCAA GGTGGCCCTT TGGCCAGCCC CTCCTTGGGT
751 CCTCCAGGAG GACTGAGCAC CCCTCCTAGC GGCATCCCTT GCCCTCCACA
801 GTGCTGCCAG GGGCAGCTCG CTCTGTGCCG TGGACTGAGA CCATCCCTG
851 GTGACAGAAT GACCCGTTTG TTGGAATGTC CTCGTTGCCA GAGAACTCC
901 CCAGGCATCT CGGAACGAAA CTATTTAGTT CCATTGTGAA CTGGCCACGG
951 GACAGCTTTT TATCAACTTA TTAAGTTGGA GCACTGTAAT CGCGCTTGCT
1001 GAGTTAGCAG TGGTGGTAAG CGTGTGTTAA ACACATAATG TTACGTTTTA
1051 GGAGAGAGAG GTCGTAAGGA AGTGTGCTGT CGCTCATGAC TCTCTCTAT
1101 TAGTTGGGTA ACAGTGGCCT CATGTTTGTG TCTGTGTGTA CACAGAGCCC
1151 TTAGGTTCTG CTCTGTTTCT TTGCCAGGTG AATGTTTGTG GCATGCGCTG
1201 CTGTCCGCGC CCCTCTGTCC TGCGCAGGGT TCAGCTGTGC GCGGCCCTGA
1251 TTCTCTCCAT GCACACAGAA CCTCCTTGTG TCTGTTCTC TGTCTCTCTG
1301 TGGCTGACTC AATAAACTTT TCCCTCTGAC ATGAAAAAAA AAAAAAAAAG
1351 AAAAAAAAAG AAAAAAAAAG AAAAAAAAAG AAAAAAAAAG
```

BLAST Results

Entry HS548358 from database EMBL:

human STS EST67250.

Score = 2126, P = 1.5e-89, identities = 444/472

Entry HS670351 from database EMBL:

human STS WI-18501.

Score = 2089, P = 7.1e-88, identities = 445/476

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 361 bp to 936 bp; peptide length: 192

Category: putative protein

Classification: no clue

```

1 MDTFSHAVSL LNFGPALATT QVRDCCCGV SLVCPASASHQ HAPLLRDTSS
51 LPPSLVPQAC REGPLLPRAP GGVLPTTWE RCQFSSELNK ARAHSMGLAQ
101 PKVLVTSSCK ASHHPPARAQ GGPLASPSLG PPGGLSTPPS GIPCPQCCQ
151 GHVALCRGLR PSPGDRMTRL LEMPRCQRNS PGISERNYLV PL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_30f4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_30f4, frame 1

Report for DKFZphtes3_30f4.1

```

[LENGTH]      192
[MW]           20281.56
[pI]           9.21
[BLOCKS]      BL01013C Oxysterol-binding protein family proteins
[KW]           All Alpha
[KW]           LOW_COMPLEXITY    10.94 %

SEQ  MDTFSHAVSLLNFGPALATTQVRDCCCGVSLVCPASASHQHAPLLRDTSSSLPPLVPQAC
SEG  .....
PRD  cccchhhheeeccccchhhhhhhccceeeccccccccccccccccccccccccccccccccc

SEQ  REGPLLPRAPGGVLPTTWERCFSSSELNKAHSMGLAQPKVLVTSSCKASHHPPARAQ
SEG  .....
PRD  cccccccccccccccccccchhhhhhhhhhhccceeeccccccccccccccccccccccccc

SEQ  GGPLASPSLPPGGLSTPPSGIPCPQCCQGHVALCRGLRPSPGDRMTRLLEMPRCQRNS
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccchhhhhhhhhccceeeccccchhhhhcccccccccc

SEQ  PGISERNYLVPL
SEG  .....
PRD  cccccccccccc

```

(No Prosite data available for DKFZphtes3_30f4.1)

(No Pfam data available for DKFZphtes3_30f4.1)

DKFZphtes3_35b4

group: cell cycle

DKFZphtes3_35b4 encodes a novel 1780 amino acid protein which is C-terminal identical to human M-phase phosphoprotein-1 (MPPL).

The novel protein contains a N-terminal Pfam kinesin motor domain and a ATP/CTP-binding site motif A (P-loop). MPPL is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.

The new protein can find application in modulation of the mitotic spindle.

"M-phase phosphoprotein-1" extension

motor protein

Sequenced by DKFZ

Locus: /map="750_H_1; 758_H_7; 759_C_9; 847_D_4; 906_D_1; 931_D_3; 944_C_1; 750_G_12; 800_A_11; 512.1 CR from top of Chr10 linkage group"

Insert length: 6284 bp

No poly A stretch found, no polyadenylation signal found

```
1 ATCGCAGTGC TGCTCGCGGG TCTGGCTAGT CAGGCGAAGT TTGCAGAATG
51 GAATCTAATT TTAATCAAGA GGGAGTACCT CGACCATCTT ATGTTTTTAG
101 TGCTGACCCA ATTGCAAGGC CTTCAGAAAT AAATTTTCGAT GGCATTAAAGC
151 TTGATCTGTC TCATGAATTT TCCTTAGTTG CTCCAAATAC TGAGGCAAAC
201 AGTTTTCGAAT CTAAAGATTA TCTCCAGGTT TGTCTTCGAA TAAGACCATT
251 TACACAGTCA GAAAAAGAAC TTGAGTCTGA GGGCTGTGTG CATATTCTGG
301 ATTCACAGAC TGTTGTGCTG AAAGAGCCTC AATGCATCCT TGGTCGGTTA
351 AGTAAAAAAA GCTCAGGGCA GATGGCACAG AAATTCAGTT TTTCCAAGGT
401 TTTTGGCCCA GCAACTACAC AGAAGGAATT CTTTCAGGGT TGCATTATGC
451 AACCACTAAA AGACCTCTTG AAAGGACAGA GTCGTCTGAT TTTTACTTAC
501 GGGCTAACCA ATTCAGGAAA AACATATACA TTTCAAGGGA CAGAAGAAAA
551 TATTGGCATT CTGCCTCGAA CTTTGAATGT ATTATTGAT AGTCTTCAAG
601 AAAGACTGTA TACAAAGATG AACCTTAAAC CACATAGATC CAGAGAATAC
651 TTAAGGTTAT CATCAGAAAC AGAGAAAGAA GAAATTGCTA GCAAAAGTGC
701 ATTGCTTCGG CAAATTAAAG AGGTACTCTG GCATATGAT AGTGATGATA
751 CTCCTTATGG AAGTTTAACT AACTCTTTGA ATATCTCAGA GTTTGAAGAA
801 TCCATAAAAG ATTATGAACA AGCCAACCTG AATATGGCTA ATAGTATAAA
851 ATTTTCTGTG TGGGTTTCTT TCTTTGAAAT TTACAATGAA TATATTTATG
901 ACTTATTGTG TCCTGTATCA TCTAAATCCC AAAAGAGAAA GATGCTGCGC
951 CTTTCCCAAG ACGTAAAGGG CTATTCTTTT ATAAAGATC TACAATGGAT
1001 TCAAGTATCT GATTCCAAAG AAGCCTATAG ACTTTTAAAA CTAGGAATAA
1051 AGCACACAGAG TGTTCCTTC ACAAAATTGA ATATGCTTC CAGTAGAAGT
1101 CACAGCATAT TCACTGTATA AATATTACAG ATTGAAGATT CTGAAATGTC
1151 TCGTGTAATT CGAGTCAGTG AATTATCTTT ATGTGATCTT GCTGGTTCAG
1201 AACGAACATAT GAAGACACAG AATGAAGGTG AAAGGTTAAG AGAGACTGGG
1251 AATATCAACA CTTCTTTATT GACTCTGGGA AAGTGTTATTA ACGTCTTGAA
1301 GAATAGTGAA AAGTCAAAGT TTCAACAGCA TGTGCCTTTC CGGGAAAGTA
1351 AACTGACTCA CTATTTTCAA AGTTTTTTTA ATGGTAAAGG GAAAAATTTGT
1401 ATGATTGTCA ATATCAGCCA ATGTTATTTA GCCTATGATG AAACACTCAA
1451 TGTATTGAAG TTCTCCGCCA TTGCACAAAA AGTTTGTGTC CCAGACACTT
1501 TAAATTCCTC TCAAGATAAA TTATTTGGAC CTGTCAAATC TTCTCAAGAT
1551 GTATCACTAG ACAGTAATTC AAACAGTAAA ATATTAAATG TAAAAAGAGC
1601 CACCATTTCA TGGGAAAATA GTCTAGAAGA TTTGATGGAA GACGAGGATT
1651 TGGTTGAGGA GCTAGAAAAC CTTGAAGAAA CTCAAATGT GGAACATAAA
1701 CTTCTTGATG AAGATCTAGA TAAACATTA GAGGAAAATA AGGCTTTTAT
1751 TAGCCACGAG GAGAAAAGAA AACTGTTGGA CTTAATAGAA GACTTGAAAA
1801 AAAAACTGAT AAATGAAAAA AAGGAAAAAT TAACCTTGGA ATTTAAAAAT
1851 CGAGAAGAAG TTACACAGGA GTTTACTCAG TATTGGGCTC AACGGGAAGC
1901 TGACTTTAAG GAGACTCTGC TTCAAGAACG AGAGATATTA GAAGAAAAATG
1951 CTGAACGTCG TTTGGCTATC TTCAAGGATT TGGTTGGTAA ATGTGACACT
2001 CGAGAAGAAG CAGCGAAAGA CATTTGTGCC ACAAAGTTG AAACGAAAGA
2051 AGCTACTGCT TGTTTGAAC TAAAGTTTAA TCAAATTTAA GCTGAATTAG
2101 CTAAACCACC AGGAGAATTA ATCAAAACCA AAGAAGGATT AAAAAAGAGA
2151 GAAATGAAT CAGATTCATT GATTCAAGAG CTTGAGACAT CTAATAAGAA
2201 AATAATTACA CAGAAATCAA GAATTAAGA ATTGATAAAT ATAATTGATC
2251 AAAAAAGAGA TACTATCAAC GAATTTTACA ACCTAAAGTC TCATATGGAA
2301 AACACATTTA AATGCAATGA CAAGGCTGAT ACATCTTCTT TAATAATAAA
2351 CAATAAATTG ATTTGTAATG AAACAGTTGA AGTACCTAAG GACAGCAAAAT
2401 CTAAAACTCTG TTCAGAAAGA AAAAGAGTAA ATGAAAATGA ACTTCAGCAA
2451 GATGAACAC CAGCAAGAA AAGGCTATAT CATGTTAGTT CAGCTATCAC
2501 TGAAGACCAA AAGAAAAGTG AAGAAGTGC ACGAACATT GCAGAAATG
2551 AAGACATCAG AGTTTACAA GAAAATAATG AAGGACTGAG AGCATTTTAA
```

2601 CTCACATTG AGAATGAAC TAAAAATGAA AAGGAAGAAA AAGCAGAATT
2651 AAATAAACAG ATTGTTCAAT TTCAGCAGGA ACTTCTCTT TCTGAAAAA
2701 AGAATTTAAC TTTAAGTAAA GAGGTCCAAC AAAATTCAGTC AAATTATGAT
2751 ATTGCAATTG CTGAATTACA TGTGCAGAAA AGTAAAAATC AAGAACAGGA
2801 GGAAAAAGATC ATGAAATTGT CAAATGAGAT AGAACTGCT ACAGAAGCA
2851 TTACAAATAA TGTTTCACAA ATAAAAATTA TGCACACGAA ATAGACGAA
2901 CTACGTACTC TTGATTCAAT TTCTCAGATT TCAACATAG ATTTGCTCAA
2951 TCTCAGGGAT CTGTCAAATG GTTCTGAGGA GGATAATTG CCAATACAC
3001 AGTTAGACCT TTTAGGTAAT GATTATTTGG TAAGTAAGCA AGTTAAAGAA
3051 TATCGAATTC AAGAACCCAA TAGGAAAAAT TCTTCCACT CTAGATTGA
3101 AGCTATTTGG GAAGAATGTA AAGAGATTGT GAAGGCTCT TCCAAAAA
3151 GTCATCAGAT TGAGGAACCT GAACAACAAA TTGAAAAAT GCAGGCAGAA
3201 GTAAAAAGCT ATAAGGATGA AAACAATAGA CTAAGAGAGA AGGAGCATAA
3251 AAACCAAGAT GACCTACTAA AAGAAAAAGA AACTCTTATA CAGCAGCTAT
3301 AAGAAGAATT GCAAGAAAAA AATGTTACTC TTGATGTTCA AATACAGCAT
3351 GTAGTTGAAG GAAAGAGAGC GCTTTCAGAA CTTACACAAG GTGTTACTTG
3401 CTATAAGGCA AAAATAAAGG AACTTGAAAC AATTTTAGAG ACTCAGAAAG
3451 TTGAACGTAG TCATTACGCC AAGTTAGAAC AAGACATTTT GGAAAGGAA
3501 TCTATCATCT TAAAGCTAGA AAGAAATTTG AAGGAATTTT AAGAATCATCT
3551 TCAGATTCTT GTCAAAACAA CCAAGATTTT AAATGTAAAG GAATCAAGC
3601 TGAAAGAGA AATCACACAG TTAACAAATA ATTTGCAAGA TATGAAACAT
3651 TTACTTCAAT TAAAGAGA AGAAGAGAAA ACCAACAGGC AAGAAACAGA
3701 AAAATTGAAA GAGGAACCTC CTGCAACCTC TGCTCGTACC CAGAATCTGA
3751 AAGCAGATCT TCAGAGGAAG GAAGAAGATT ATGCTGACCT GAAGAGAGAA
3801 CTGCTGATG CCAAAAGCA GATTAGCAA GTACAGAAAG AGGTATCTGT
3851 AATCGGTGAT GAGGATAAAT TACTGAGGAT TAAATTAAT GAATCGGAGA
3901 AAAAGAAAAA CCAGTGTCT CAGGAATTAG ATATGAAGCA GCCAACCAT
3951 CAGCAACTCA AGGAGCAGTT AAATATCAG AAGTGGAG AAGCTATACA
4001 ACAGTATGAG AGAGCATGCA AAGATCTAAA TGTAAAGAG AAAATAATTG
4051 AAGACATGCG AATGACACTA GAAGAACAGG AACAACTCA GTAGAACAG
4101 GATCAAGTGC TTGAGGCTAA ATTAGAGGAA GTTGAAGGC TGCCACAGA
4151 ATTTGAAAAA TGGAGGAAA AATGCAATGA TTTGGAACC AAAAACATC
4201 AAAGGTCAA TAAAGACAT GAGAACAACA CAGATGTGCT TGGAAAGCTC
4251 ACTAATCTTC AAGATGAGTT ACAGGAGTCT GAACAGAAAT ATAATGCTGA
4301 TAGAAGAAA TGGTTAGAAG AAAAATGAT GCTTATCACT CAAGCGAAG
4351 AAGCAGAGAA TATACGAAAT AAAGAGATGA AAAAATATGC TGAGGACAGG
4401 GAGCGTTTTT TTAAGCAACA GAATGAAATG GAAATACTGA CAGCCAGCT
4451 GACAGAGAAA GATAGTGACC TTCAAAAGTG GCGAGAGAA CGAGATCAAC
4501 TGGTTGACG TTTAGAAATA CAGCTAAAG CACTGATATC CAGTAATGTA
4551 CAGAAAGATA ATGAAATTGA ACACTAAAA AGGATCATAT CAGAGACTTC
4601 TAAATAGAAA ACACAAATCA TGGATATCAA GCCCAACGT ATTAGTTGAG
4651 CAGATCTGTA CAACTTCAA ACTGAACCTC TATCGACAAG TTTTGAAAT
4701 TCCAGAAATA AAATAGAGGA TGGATCTGTA GTCCTTGACT CTTGTGAGT
4751 GTCAACAGAA AATGATCAA GCATCTGATT TCCAAAACCT GAGTTAGAGA
4801 TTCAATTTAC ACCTTTACAG CCAACAAAAA TGGCAGTGAA ACACCTGGT
4851 TGTACCACAC CAGTGACAGT TGAGATTCCC AAGGCTCGGA AGAGGAAGAG
4901 TAATGAAATG GAGGAGGACT TGGTGAAATG TGAAAAAAG AAGAATGCTA
4951 CACCAGAAC TAATTGAAA TTCTCTATT CAGATGATAG AAATCTTCT
5001 GTCAAAAAGG AACAAAAGGT TGCCATACGT CCATCATCTA AGAAACATA
5051 TTCTTTACGG AGTCAGGCAT CCATAATTGG TGTAAACCTG GCCACTAAGA
5101 AAAAAGAAAG AACACTACAG AAATTTGGAG ACTTCTTACA ACATTCTCCC
5151 TCAATTTCTT AATCAAAAGC AAAGAAGATA ATTGAAACAA TGAGCTCTTC
5201 AAAGCTCTCA AATGTAGAAG CAAGTAAAGA AAATGTGTCT CAACCAAAAC
5251 GAGCCAAACG GAATTTATAC ACAAGTGAAA TTTCACTCC TATTGATATA
5301 TCAGGCCAAG TGATTTTAAT GGACCAGAAA ATGAAGGAGA GTGATCACC
5351 GATTATCAAA CGACGACTTC GAACAAAAC AGCCAATAA ATCACTTATG
5401 GAAATGTTTA ATATAAATTT TATAGTCATA GTCATTGGAA CTTGCATCT
5451 GTATTGTAAA TATAAATGTA TATATTATGC ATTAATCAC TCTGCATATA
5501 GATTGCTGTT TTATACATAG TATAATTTA ATTCATAAA TGAGTCAAAA
5551 TTTGTATATT TTTATAAGGC TTTTATATA TAGCTTCTT CAACTGTAT
5601 TTCCCTATTA TCTCAGACAT TGGATCAGTG AAGATCCTAG GAAAGAGGCT
5651 GTTATTCTCA TTTATTTTGC TATACAGGAT GTAATAGGT AGGTATTTGG
5701 TTTACTTATA TTTAACAATG TCTTATGAAT TTTTCTACT TTATCTGTTA
5751 TACAACGTAT TTTACATATC TGTTTGGATT ATAGCTAGGA TTTGGAGAA
5801 AAGTGTGTAC AGATCAGAAA ACATGTATAT ACATTATTTA GAAAAGATCT
5851 CAAGTCTTTA ATTAGAATGT CTCATTATT TTGTAACAT TTTGTGGGTA
5901 CATAGTACAT GTATATATT ACGGGGTATG TGAGATGTTT TGACACAGGC
5951 ATGCAATGTG AAATACGTGT ATCATGGAGA ATGAGGTATC CATCCCTCA
6001 AGCATTTTTT CTTTGAATTA CAGATAATCC AATTACATT TTTAGATCAT
6051 TAAAAAATAT ACAAGTAAGT TATTATTGAT TATAGTCACT CTATTGTGCT
6101 ATCAGATAGT AGATCATTCT TTTTATCTTA TTTGTTTTG TACCAATTAA
6151 CCATCCCCAC CTCCCTGTC AACCGTCAGT ACCCTTACCA GCCACTGGTA
6201 ACCATTCTTC TACTCTGTAT GCCCATGAG TCAATTGATT TTATTTTATG
6251 ATCCCATAAA TAAATGAGAA CATGCAAAAA AAAA

BLAST Results

Entry HS898149 from database EMBL:
human STS WI-9217.

Score = 4247, P = 1.5e-187, identities = 855/862

Medline entries

94119956:
Cloning of cDNAs for M-phase phosphoproteins recognized
by the MPM2 monoclonal antibody and determination of the
phosphorylated epitope.

98101856:
Interaction of a Golgi-associated kinesin-like protein with
Rab6.

95122643:
Identification and partial characterization of mitotic
centromere-associated kinesin, a
kinesin-related protein that associates with centromeres during
mitosis.

Peptide information for frame 3

ORF from 48 bp to 5387 bp; peptide length: 1780
Category: known protein
Classification: Cell structure/motility
Prosite motifs: ATP_GTP_A (152-160)

```
1 MESNFNQEGV PRPSYVFSAD PIARPSEINF DGIKLDLSHE FSLVAPNTEA
51 NSFESKDYLO VCLRIRPFTQ SEKELESEGC VHILDSQTVV LKEPQCILGR
101 LSEKSSGQMA QKFSFSKVFG PATTQKEFFQ GCIMQPVKDL LKGQSRILFT
151 YGLTNSGKTY TFQGTENIG ILPRTLNVLF DSLQERLYTK MNLKPHRSRE
201 YLRSSSEQEK EEIASKSALL RQIKVTVHN DSDDTLYGSL TNSLNISEFE
251 ESIKDYEQAN LNMANSIKFS VVVSFFFIYN EYIYDLFVPV SSKFQKRKML
301 RLSQDVKGYS FIKDLQWIVQ SDSKEAYRLL KLGIKHQSVA FTKLNNASSR
351 SHSIFTVKIL QIEDSEMSRV IRVSELSLCD LAGSERTMKT QNEGERLRET
401 GNINTSLTLL GKCINVLKNS ESKSFQOHVP FRESKLTHYF QSFFNGKGKI
451 CMIVNISQCY LAYDETLNVL KFSIAIAQKVC VPDTLNSSQD KLFQPVKSSQ
501 DVSLDSNSNS KILNVKRATI SWENSLEDLM EDEDLVEELE NAEETQNVET
551 KLLDEDLDKT LEENKAFISH EKKRLLDLI EDLKKKLINE KKEKLTLEFK
601 IREEVTQEFQ QYWAQREADF KETLLQEREI LEENAERRLA IFKDLVGKCD
651 TREEAARDIC ATKVETEEAT ACLELKFNOI KAEAKTKGE LIKTKEELKK
701 RENESDSLII ELETSSNKII TQNQRILKEL NIIDQKEDTI NEFQNLKSHM
751 ENTFCNDKA DTSSLIINNK LICNETVEVP KDSKSKICSE RKRNVENELQ
801 QDEPPAKKGS IHVSSAITED QKKSEEVVPN IAEIEDIRVL QENNEGLRAF
851 LLTINELKN EKEEKAELNK QIVHFQOELS LSEKKNLTLN KEVQIQSINY
901 DIAIAELHVQ KSKNQEQEEK IMKLSNEIET ATRSIITNVS QIKLMHTKID
951 ELRTLDSVSQ ISNIDLLNLR DLSNGSEEDN LPNTQLDLLG NDYLVSKQVK
1001 EYRIQEPNRE NSFHSSIEAI WEECKEIVKA SSKKSHQIEE LEQQIEKLQA
1051 EVKGYKDENN RLKEKEHKNQ DDLLEKETL IQQLKEELQE KNVTLQVQIO
1101 HVVEGKRALS ELTQGVTCYK AKIKELETIL ETQKVERSIS AKLEQDILEK
1151 ESIIKLERN LKEFQERLQD SVKNTKDLNV KELKLKEEIT QLTNNLQDMK
1201 HLLQLKEEEE ETNRQETEKL KEELSASSAR TQNLKADLQR KEEDYADLKE
1251 KLTDAAKQIK QVQKEVSVMR DEDKLLRIKI NELEKKKNQC SQELDMKQRT
1301 IQQLKEQLNN QKVEEAIQY ERACKDLNVK EKIIDMRMT LEEQEQTQVE
1351 QDQVLEAKLE EVERLATELE KWKEKCNLE TKNNQRSNKE HENNTDVLGK
1401 LTNLQDELQE SEQKYNADRK KWLEEKMLI TQAKEAENIR NKEMKKYAE
1451 RERFFKQONE MEILTAQLTE KSDSLQKWE ERDQLVALE IQLKALISSN
1501 VOKDNEIEQL KRIISETSKI ETQIMDIKPK RISSADPKL QTEPLSTSFE
1551 TSRNKIEGDS VVLDSCEVST ENDQSTRFPK PELEIQFTPL QPNKMAVKHP
1601 GCTTPVTVEI PKARKRSNE MEEDLVKCN KKNATPRTNL KFPISDRNS
1651 SVKKEQKVAI RPSSKKTYSL RSQASIIQVN LATKKKEGTL QKFGDFLOHS
1701 PSILQSKAKK IETMSSSKL SNVEASKENV SQPKRAKRL YLSEISSPID
1751 ISGQVILMDQ KMKESDHQII KRLRTKTAK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b4, frame 3

TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase
phosphoprotein-1 mRNA, partial cds., N = 1, Score = 3743, P = 0

PIR:A36881 MPM2-reactive phosphoprotein 1 - human (fragment), N = 2,
Score = 2808, P = 2.5e-294

TREMBL:AF070672_1 product: "rabkinesin6"; Homo sapiens rabkinesin6
mRNA, complete Cds., N = 2, Score = 680, P = 2.6e-99

>TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase
phosphoprotein-1 mRNA, partial cds.
Length = 753

HSPs:

Score = 3743 (561.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 752/753 (99%), Positives = 753/753 (100%)

```
Query: 1028 VKASSKSHQIEELEQQIEKLQAEVKGKDYNNRLKEKEHKNQDDLLKEKETLIQQLKEE 1087
          VKASSKSHQIEELEQQIEKLQAEVKGKDYNNRLKEKEHKNQDDLLKEKETLIQQLKEE
Sbjct: 1 VKASSKSHQIEELEQQIEKLQAEVKGKDYNNRLKEKEHKNQDDLLKEKETLIQQLKEE 60

Query: 1088 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
          LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI
Sbjct: 61 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 120

Query: 1148 LEKESIIKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 1207
          LEKESIIKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE
Sbjct: 121 LEKESIIKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 180

Query: 1208 EEEETNRQETEKLEELSSASARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 1267
          EEEETNRQETEKLEELSSASARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS
Sbjct: 181 EEEETNRQETEKLEELSSASARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 240

Query: 1268 VMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL 1327
          VMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL
Sbjct: 241 VMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL 300

Query: 1328 NVKEKIIEDMRMTLEEQEQTVQEQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS 1387
          NVKEKIIEDMRMTLEEQEQTVQEQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS
Sbjct: 301 NVKEKIIEDMRMTLEEQEQTVQEQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS 360

Query: 1388 NKEHENNTDVLGKLTNLQDELQSEQKYNADRKWLEKMMMLITQAKEAENIRNKEMKKY 1447
          NKEHENNTDVLGKLTNLQDELQSEQKYNADRKWLEKMMMLITQAKEAENIRNKEMKKY
Sbjct: 361 NKEHENNTDVLGKLTNLQDELQSEQKYNADRKWLEKMMMLITQAKEAENIRNKEMKKY 420

Query: 1448 AEDRERFFKQONEMEILTAQLTEKSDSLQKWREERDQVAALEIQLKALISSNVQKDNEI 1507
          AEDRERFFKQONEMEILTAQLTEKSDSLQKWREERDQVAALEIQLKALISSNVQKDNEI
Sbjct: 421 AEDRERFFKQONEMEILTAQLTEKSDSLQKWREERDQVAALEIQLKALISSNVQKDNEI 480

Query: 1508 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESRNKIEDGSSVVLDSCE 1567
          EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESRNKIEDGSSVVLDSCE
Sbjct: 481 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESRNKIEDGSSVVLDSCE 540

Query: 1568 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKRKSNEEEDLVK 1627
          VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKRKSNEEEDLVK
Sbjct: 541 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKRKSNEEEDLVK 600

Query: 1628 CENKKNATPRTNLKFPIISDRNSSVKKEQKVAIRPSSSKTYSLSQASIIGVNLATKKKE 1687
          CENKKNATPRTNLKFPIISDRNSSVKKEQKVAIRPSSSKTYSLSQASIIGVNLATKKKE
Sbjct: 601 CENKKNATPRTNLKFPIISDRNSSVKKEQKVAIRPSSSKTYSLSQASIIGVNLATKKKE 660

Query: 1688 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 1747
          GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS
Sbjct: 661 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 720

Query: 1748 PIDISGQVILMDQMKESDHQIIKRLRLTKTAK 1780
          PIDISGQVILMDQMKESDHQIIKRLRLTKTAK
Sbjct: 721 PIDISGQVILMDQMKESDHQIIKRLRLTKTAK 753

Score = 197 (29.6 bits), Expect = 2.1e-11, P = 2.1e-11
Identities = 114/542 (21%), Positives = 253/542 (46%)

Query: 692 IKTKLELKKRENESDSLIQELETSSNKKIITQNRKELINIIDQKEDTINEFQNLKSHM- 750
          +K + + E + I++L+ K +N R+KE + ++D + E + L +
Sbjct: 1 VKASSKSHQIEELEQQIEKLQAEVKGKDYNNRLKEKEHKNQDDLLKEKETLIQQLK 58

Query: 751 ENTFCNKDADTS-SLIINNKLICNETVEVPKDSKSKICSERKRVNENELQDDEPPAK-- 807
          E + N D ++ K +E + K+KI E + + E + + AK
Sbjct: 59 EELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKI-KELETILETQKVERSHSAKLE 117

Query: 808 KGSIHVSSAITEDQKSEEVPRNIAE-IEDIRVLQENNEGLRAFLLLTIENELKNEK---- 862
```


+ + S I + ++ +E + ++ + ++ + L L+ + + N L++ K
 Sbjct: 118 QDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKKEEITQLTNNLQDMKHLLO 177
 Query: 863 --EEKAELNKQIVH-FQOELSLSEKKNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEE 919
 EE+ E N+Q ++ELS S + L ++Q+ + +Y A+L K K + ++
 Sbjct: 178 LKEEEETNRQETEKLEELSASSARTQNLKADLQKKEEDY---ADL---KEKLTDAKK 230
 Query: 920 KIMKLSNEIETATRSITNNVSIQIKLMHTKIDEL-RTLDSVSIQISNIDLLNLRDLSNGSEE 978
 +I ++ E+ S+ + + KL+ KI+EL + + SQ +D+ R + E+
 Sbjct: 231 QIKQVQKEV-----SVMRD--EDKLLRIKINELEKKKNQCSQ--ELDMKQ-RTIQQKKEQ 280
 Query: 979 DNLPTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAIWEECKEIVKASSKSHQI 1038
 N N +++ Y + K+ ++E E+ ++E + E + K +++
 Sbjct: 281 LN--NQKVEEAIQQY--ERACKDLNVKEKIID--MRMTLEEQQEQTVQEQDQVLEAKLEEV 335
 Query: 1039 EEEQQIEKLQAEVKGKYNRNLKEKEHKNQDDLLKEKETLIQQLKEELQEKNTV--- 1094
 E L ++EK + + + NN+ KEH+N D+L + L +L+E Q+ N
 Sbjct: 336 ERLATELEKWKKECNDLETNNQKRSNKEHENNTDVLGKLTNLQDELQEQSKYNADKKW 395
 Query: 1095 LDVQIQHVVEGKRA-----LSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
 L+ ++ + + K A + + + + + + E+E IL Q E+ + ++
 Sbjct: 396 LEEKMMLITQAKEAENIRNKEMKKAEDRERFFKQKNEME-ILTAQLTEKDSDLQKWRE- 453
 Query: 1148 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELK-LKEEITQLTNNLQDMKHLQLK 1206
 E++ ++ LE LK + +V+ KD ++LK + E +++ + D+K +
 Sbjct: 454 -ERDQLVALEIQLKAL---ISSNVQ--KDNEIEQLKRIISETSKIETQIMDIK---PKR 504
 Query: 1207 EEEETNRQETEKLEELSASSARTQN 1233
 + ++ +TE L S + ++
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIED 531
 Score = 186 (27.9 bits), Expect = 3.2e-10, P = 3.2e-10
 Identities = 131/674 (19%), Positives = 294/674 (43%)
 Query: 673 LELKFNQIKAEAKTGLIKT-KEELKKRENSDSLIQELTSNKKIITQNRIKELIN 731
 L+ K ++ + +L K K LI+ KEEL+++ D IQ + + + Q +
 Sbjct: 35 LKEKEHKNQDDLLKEKETLIQQLKEELQEKNTLDVQIQHVVEGKRALSELTQGVTCYKA 94
 Query: 732 IIDQKEDTINEFQNL-KSHMENTFKCNDKADTSSLIINNKLICNETVEVPKDSKSKICSE 790
 I + E TI E Q + +SH + D + S+I+ + E E +DS
 Sbjct: 95 KIKELE-TILETQKVERSHSAKLEQ--DILEKESIILKLERNLKEFQEHLDQDS---VKN 147
 Query: 791 RKRNVENELQ-QDEPPAKKGSIHVSSAITEDQKKSEEV-RPNIAEI-EDIRVLQENNEGL 847
 K +N EL+ ++E ++ + + ++ EE R ++ E++ + L
 Sbjct: 148 TDLNVLKELKKEEITQLTNNLQDMKHLQLKKEEEETNRQETEKLEELSASSARTQNL 207
 Query: 848 RAFLTIENELKNEKEEKAELNKQIVHFQOELSLSEKKNLTLSKEVQQI-----QSNYDI 902
 +A L E + + KE+ + KQI Q+E+S+ ++ L ++ ++ Q + ++
 Sbjct: 208 KADLQKKEEDYADLKEKLTDAKKQIKQVQKEVSVMRDEKLLRIKINELEKKKNQCSQEL 267
 Query: 903 AIAELHVQKSKNQEQEEKIMKLSNEIETATRSITNNVSIQIKLMHTKIDEL-RTLDSVSIQ 961
 + + +Q+ K Q +K+ + + E A + + I+ M ++E +T Q+
 Sbjct: 268 DMKQRTIQQKLEQLNNQKVEEAIQQYERACKDLNVKEKIIDMRMTLEEQQEQTVQEQDQV 327
 Query: 962 SNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRI--QEPNRENSFHSSIEA 1019
 L + L+ E+ L+ N + + + N ++ S +
 Sbjct: 328 LEAKLEEVEERLATELEKWKKECNDLETNNQKRSNKEHENNTDVLGKLTNLQDELQEQSK 387
 Query: 1020 IWEECKEIVKASSKSHQIEELEQQIEKLQAEVKGKYNRNLKEKEHKNQ--DILLKEK 1077
 + K+ ++ Q+E E K E+K Y ++ R +++++ + L EK
 Sbjct: 388 YNADRKKWLEEKMLITQAKEAENIRNK---EMKKAEDRERFFKQKNEMEILTAQLTEK 444
 Query: 1078 ETLIQQLKEELQEKNTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVER 1137
 ++ +Q+ +EE + L++Q++ ++ + + ++ ++ET + K +R
 Sbjct: 445 DSDLQKWREERDQLVALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKR 504
 Query: 1138 SHSAKLEQDILEKESIILKLERNLKEFQEHLDQDS---VKNTKDLNVKELKKEEITQLT 1193
 SA ++ E S ++ RN E + DS +N + + +L+ + T L
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIEDGSVVLDSCVSTENDQSTRFPKPELEIQFTPLQ 564
 Query: 1194 NNLQDMKH---LLQKKEEEETNRQETEKLEEL-SASSARTQNLKADLQKKEEDYADLK 1249
 N +KH + + + ++++++E+L + + + +L+ D +
 Sbjct: 565 PNKMAVKHPGCTTPVTVKIPKARKKRSNEMEEDLVKCNKKNATPRTNLKFPISSDDRNSS 624
 Query: 1250 EKLTDAKKQIKQVQKEVSVMRDEKLLRIKINELEKKKNQCSQEL-DMKQRTIQQKLEQL 1308
 K + K I+ K+ +R + + I +N KKK Q+ D Q + L+ +
 Sbjct: 625 VK-KEQKVAIRPSSKKTYSLSQASI--IGVNLATKKEGTLQKFGDFLQHSFSLQSKA 681
 Query: 1309 NNQKVEEAIQQYERACKDLNVKEKIIDMR 1338
 +K+ E + + + + + KE + + R
 Sbjct: 682 --KKIETMSSSKLSNVEAS-KENVSQPKR 708

Score = 165 (24.8 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 140/626 (22%), Positives = 271/626 (43%)

Query: 536 VEELNAEETQNVETKLLDEDLTKLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEK- 594
+EELE E E K +D + L+E + H+ + LL E L +L E +EK
Sbjct: 11 IEELQEQIEKQAEVKGY-KDENNRLEKE---HKNQDDLLKEKETLIQQLKEELQEK 65

Query: 595 LTLEFKIREEVT-----QFTQYWAQREADFKE--TLQEREILEENAERRLAIFKDLVG 647
+TL+ +I+ V E TQ +A KE T+L+ +++ E + +L +D++
Sbjct: 66 VTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV-ERSHSKLE--QDILE 122

Query: 648 KCDT---REEAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENE 704
K E K+ ++ + T L +K +K E+ + L K L+ +E E
Sbjct: 123 KESIILKLERNLKEFEHLQDSVKNTKDLNVKELKLEETQLTNNLQDMKHLQLKEE 182

Query: 705 SDSLIQELETSSNKKIITONQRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSS 764
++ QE E +++ + R + L + +KE+ + + + K K + S
Sbjct: 183 EETNRQETKLEKELSSASSARTONLQADLQKEEDYADLKEKLTDAKKQIKQVQK-EVSV 241

Query: 765 LIINNKLICNETVEVPKDSKSKICSEKRVNENELQDEPPAKKSGSIHVSSAITEDQKKS 824
+ +KL+ + E+ K K CS+ + + +QQ + V AI + ++
Sbjct: 242 MRDEKLLRIKINELEK--KKNQCSQELDMKQRTIQQLEQLNNQK--VEEAIQYERAC 297

Query: 825 EEVPRNIAEIEDIRVLQENNEGLRAFLTIENELKNEKEEAKLNKQIVHFQQLSLSEK 884
+++ IED+R+ E E + + + L+ + EE L ++ +++ + E
Sbjct: 298 KDLNVKEKIIEDMRMTLEEQQEQ---VEQDQVLEAKLEEVEERLATELEKWEKCNDET 354

Query: 885 KNLTLSEVQIQSNYDIAIAELHVQSKNQEQEEKIMKLSNE-IETATRSITN-----N 938
KN S + + ++N D+ + +L + + QE E+K + +E IT N
Sbjct: 355 KNNQRSNK--EHEENTDV-LGKLTNLQDELQSEQKYNADRKWLEEKMLITQAKEAEN 411

Query: 939 VSQIKLMHTKIDELRTLDVSQISNIDL-LNLRD--LSNGSEEDNLPNTQLDLLGNDYLV 995
+ + D R +++ + L +D L EE + L++ +
Sbjct: 412 IRNKEMKKAEDRERFFKQKNEMEILTAQLTEKSDQKWRERDQLVAALEIQLKALIS 471

Query: 996 SKQVKYRIQEPNRENSFHSSIEA-IWE-ECKEIVKASSKKSHQIEELEQIEKQAEVK 1053
S K+ I++ R S S IE I + + K I A K Q E L E + +++
Sbjct: 472 SNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKL-QTEPLSTSFEISRNKIE 530

Query: 1054 GYKDNNRLKEKEHKNQDDLLKEKE-----TLIQQLEKEELQEKNVTLVDVQIQHVVEGKRA 1108
+ + +Q + E T +Q K ++ T V ++ KR
Sbjct: 531 DGSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVHKPGCTPTVTKIPKARRK 590

Query: 1109 LSELTQG-VTCYKAKIKELETILETQ-KVERSHSAKLEQDILEKES 1152
+E+ + V C K T L+ +R+ S K EQ + + S
Sbjct: 591 SNEMEEDLVKCNKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPS 636

Score = 143 (21.5 bits), Expect = 1.3e-05, P = 1.3e-05
Identities = 164/684 (23%), Positives = 304/684 (44%)

Query: 295 QRRKMLR-LSQDVKGYSFIKDLQWIOVSDSKEAYRLKLGIKHQSVAFTKLNNASS---- 349
+K +++ L +++ + D+Q V + K A L G+ +L
Sbjct: 49 EKETLIQQLKEELQEKNVTLVDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV 108

Query: 350 -RSHSI-FTVKILQIEDSEMSRVIRVSELSLCLAGSERTMKTONEGE-RLRETGNINTS 406
RSHS IL+ E + + E L S + K N E +L+E T+
Sbjct: 109 ERSHSAKLEQDILEKESIIKLERNLKEFQE-HLQDSVKNTKDLNVKELKLEETQLTN 167

Query: 407 LLTLGKCINVLNSEKSKFQHVFPRESKLTHYFQSFFNGKGKICMIVNISQCYLAYDET 466
L K + LK E+ +Q + +L+ N K + + Y E
Sbjct: 168 NLQDMKHLQLKEEEEETNRQETKLEKELSSASSARTONLQADL---QRKEEDYADLKEK 224

Query: 467 LNVLFKSAIAQKVCVPTLNSSQDKLFGPVKSSQDVSLSNSNSKILNVKRATISWENSL 526
L K I Q V ++ +DKL +K ++ + N S+ L++K+ TI
Sbjct: 225 LTDAK-KQIKQ-VQKEVSMRDEKLLR-IKINE-LEKKKNQCSQELDMKQRTIQQLEKQ 280

Query: 527 EDLMEEDLVEELNAEETQNVETKLLDEDLTKLEENKAFISHEEKRKLLDL-IEDLKK 585
+ + E+ +++ E A + NV+ K++ ED+ TLEE + + E+ ++L+ +E+++
Sbjct: 281 LNNQVVEEAIQYERACKDLNVKEKII-EDMRMTLEEQQEQ--TQVEQDQVLEAKLEEVE 337

Query: 586 KLIN-EK-KEKLT-LEFKIREEVTQFTQYWAQREADFKETLLQEREILEE---NAERR 638
EK KEK LE K + +E + K T LQ+ E+ E NA+R+
Sbjct: 338 LATELEKWEKCNDETCKNNQRSNKEHEN---NTDVLGKLTNLQD-ELQSEQKYNADRK 393

Query: 639 LAIFKDLVKGCDTREAADKICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEEL 698
+ + ++ T+ + A++I K E ++ E F Q + E+ +L + +L
Sbjct: 394 KWLEEKMM--LITQAKEAENI-RNK-EMKKAEDRERFFKQ-QNEMEILTAQLTEKSDL 448

Query: 699 KKRENSDSLIQELETSSNKKIITON-QR---IKELINIIDQKEDTINEFQNLKSHMENTF 754
+K E D L+ LE K +I+ N Q+ I++L II + + ++K ++

Sbjct: 449 QKWEERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSA 508

Query: 755 KCNDKADTSSLIINNKLICN--ETVEVPKDSKSKICSERK---RVNENELQ-QDEP--PA 806
DK T L + ++ N E V DS ++ +E R + EL+ Q P P

Sbjct: 509 D-PDKLQTEPLSTSFEISRNKIEDGSVVLDS-CEVSTENDQSTRFPKPELEIQFTPLQPN 566

Query: 807 KKGSIH--VSSAITEQKKSEEVPRNIAEIEDIRVLQENNEGLRA---FLLTIENELKNE 861
K H ++ +T K+ + + N E + ++ + N R F ++ + +

Sbjct: 567 KMAVHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKNATPRTNLKFPISDDRNSSVK 626

Query: 862 KEEKAEL---NKQIVHFQOELSLSSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQE 918
KE+K + +K+ + + S+ NL K+ +Q D + +SK ++

Sbjct: 627 KEQKVATRPSSKKTYSLSQASIIGV-NLATKKKEGTLQKFGDFLQHSPSILQSKAKKII 685

Query: 919 EKIM--KLSNEIETATRSITNNVSIQKLMHTKI--DELRT-LDSVSQISNID 965
E + KLSN +E + NVSQ K + K+ E+ + +D Q+ +D

Sbjct: 686 ETMSSSKLSN-VEASKE---NVSQPKRAKRLTYSEISSPIDISGQVILMD 732

Score = 133 (20.0 bits), Expect = 1.6e-04, P = 1.6e-04
Identities = 94/426 (22%), Positives = 188/426 (44%)

Query: 527 EDLM-EDDELVEELENAEETQNVETKLLDEDLDTLEENKAFISHEEKRLLOL-IEDLK 584
+DL+ E E L+++L+ + +NV LD + +E +A + I++L+

Sbjct: 44 DDLKKEKETLIQQLKEELQEKNV---LDVQIQHVVEGKRALSELGTQGVTCYKAKIKELE 100

Query: 585 KKLINKEKELTLEFKIREEVQ-EFTQYWAQREA-DFKETLLQEREILEENAERRLAIF 642
L +K E+ + K+ +++ + E +R +F+E L + ++ + L +

Sbjct: 101 TILETQKVER-SHSAKLEQDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKL- 158

Query: 643 KDLVGKCDTREEAAKIDICATKVETEATACLELKFNQIAELAKTKGELIKTKEELKKRE 702
K+ + + + K+ K E EE + ++K EL+ + K +L+++E

Sbjct: 159 KEETITQLTNNLQDMKHLQLKEEEETN---RQTEKLKEELSASSARTQNLKADLQKE 215

Query: 703 NESDSLQIELETSNKKIITQNRKIKELINIIDQK-EDTINEFQNLKSHMENTFKCNDKA- 760
+ L ++L T KK I Q Q+ ++ D+ INE + K+ +

Sbjct: 216 EDYADLKEKL-TDAKKQIKQVQKEVSMRDEKLLRLIKINELEKKKQCSQELDMKQRTI 274

Query: 761 DTSSLIINNKLICNETVE---VPKDS--KSKICSE-RKRVNENE---LQDEPPAKKGS 810
+NN+ + E ++ KD K KI + R + E E ++QD+ K

Sbjct: 275 QQLKEQLNNQKV-EBAIQYERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLE 333

Query: 811 IHVSSAITEQKKSEEVPR-NIAEIEDIRVLQENNEGLRAFLTTIENELKNEKEKAELN 869
V TE +K E+ + ENN + L +++EL+ E E+K +

Sbjct: 334 -EVERLATELEKWEKCNLDLETKNQSRNKEHENNTDVLGKLTNLQDELQ-ESEQYNAD 391

Query: 870 KQIVHFQOELSLSSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNEIE 929
++ ++++ L +T +KE + I++ + K E E+ K NE+E

Sbjct: 392 RK-KWLEEKMMML-----ITQAKEAENIRNK-----EMKKYAEDRRERFFKQONEME 435

Query: 930 TATRSITNNVSIQKLMHTKIDEL 952
T +T S ++ + D+L

Sbjct: 436 ILTAQLTEKDSDLQKWEERDQL 458

Pedant information for DKFzptes3_35b4, frame 3

Report for DKFzptes3_35b4.3

[LENGTH] 1780
[MW] 206176.77
[pI] 5.60
[HOMOL] TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds. 0.0
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-21

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 6e-20
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w
 MYO1 - myosin-1 isoform] 4e-19
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 1e-15
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 jannaschii, MJ1322] 2e-14
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-09
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YKL179c] 3e-09
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 2e-07
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-07
 [FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c]
 2e-07
 [FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 1e-06
 [FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
 3e-06
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YDR217c] 4e-06
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 2e-05
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YAL035w] 2e-04
 [FUNCAT] r general function prediction [M. jannaschii, MJ1254] 0.001
 [BLOCKS] BL00387A
 [BLOCKS] BL00411H
 [BLOCKS] BL00411G
 [BLOCKS] BL00411F
 [BLOCKS] BL00411E Kinesin motor domain proteins
 [BLOCKS] BL00411D Kinesin motor domain proteins
 [BLOCKS] BL00411C Kinesin motor domain proteins
 [BLOCKS] BL00411B Kinesin motor domain proteins
 [BLOCKS] BL00411A Kinesin motor domain proteins
 [SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus)] 2e-68
 [SCOP] d2tmab 1.105.4.1.1 Tropomyosin [rabbit (Oryctolagus cuniculus)] 4e-05
 [SCOP] d3kar 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 2e-09
 [EC] 3.6.1.32 Myosin ATPase 5e-25
 [PIRKEW] nucleus 4e-27
 [PIRKEW] phosphotransferase 3e-16
 [PIRKEW] duplication 6e-20
 [PIRKEW] citrulline 6e-18
 [PIRKEW] tandem repeat 4e-24
 [PIRKEW] heterodimer 3e-28
 [PIRKEW] endocytosis 1e-23
 [PIRKEW] heart 1e-17
 [PIRKEW] transmembrane protein 2e-28
 [PIRKEW] serine/threonine-specific protein kinase 3e-16
 [PIRKEW] zinc finger 1e-23
 [PIRKEW] surface antigen 2e-16
 [PIRKEW] DNA binding 1e-25
 [PIRKEW] metal binding 1e-23
 [PIRKEW] muscle contraction 4e-24
 [PIRKEW] heterotetramer 4e-24
 [PIRKEW] acetylated amino end 2e-19
 [PIRKEW] actin binding 5e-25
 [PIRKEW] mitosis 3e-58
 [PIRKEW] microtubule binding 3e-58
 [PIRKEW] ATP 3e-58
 [PIRKEW] thick filament 4e-24
 [PIRKEW] phosphoprotein 9e-29
 [PIRKEW] leucine zipper 1e-12
 [PIRKEW] skeletal muscle 8e-24
 [PIRKEW] disulfide bond 1e-12
 [PIRKEW] heterotrimer 1e-29
 [PIRKEW] calcium binding 6e-18
 [PIRKEW] alternative splicing 4e-21
 [PIRKEW] P-loop 2e-63
 [PIRKEW] coiled coil 3e-58
 [PIRKEW] heptad repeat 1e-25
 [PIRKEW] methylated amino acid 4e-24
 [PIRKEW] peripheral membrane protein 1e-23
 [PIRKEW] dimer 1e-12
 [PIRKEW] cardiac muscle 1e-17
 [PIRKEW] hydrolase 5e-25
 [PIRKEW] microtubule 6e-15
 [PIRKEW] muscle 7e-23
 [PIRKEW] membrane protein 6e-20
 [PIRKEW] GTP binding 8e-22
 [PIRKEW] EF hand 6e-18
 [PIRKEW] cell division 1e-25
 [PIRKEW] cytoskeleton 4e-24
 [PIRKEW] hair 6e-18
 [PIRKEW] Golgi apparatus 8e-24
 [PIRKEW] calmodulin binding 1e-23

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 3e-16
 [SUPFAM] myosin motor domain homology 5e-25
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-13
 [SUPFAM] kinesin-related protein KIP1 9e-27
 [SUPFAM] kinesin-related protein CIN8 4e-36
 [SUPFAM] kinesin heavy chain 4e-24
 [SUPFAM] plectin 1e-13
 [SUPFAM] trichohyalin 6e-18
 [SUPFAM] kinesin-related protein KIF3 1e-29
 [SUPFAM] kinesin-related protein KIF2 3e-20
 [SUPFAM] ribosomal protein S10 homology 1e-13
 [SUPFAM] giantin 8e-24
 [SUPFAM] protein kinase homology 3e-16
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-13
 [SUPFAM] kinesin-related protein unc-104 8e-26
 [SUPFAM] human early endosome antigen 1 1e-23
 [SUPFAM] unassigned kinesin-related proteins 1e-28
 [SUPFAM] Mycoplasma genitalium hypothetical protein MG218 4e-17
 [SUPFAM] myosin heavy chain 5e-25
 [SUPFAM] conserved hypothetical P115 protein 4e-20
 [SUPFAM] centromere protein E 5e-24
 [SUPFAM] calmodulin repeat homology 6e-18
 [SUPFAM] kinesin-related protein KLP61F 1e-25
 [SUPFAM] hypothetical protein MJ0914 3e-12
 [SUPFAM] kinesin-related protein MKLP-1 2e-63
 [SUPFAM] pleckstrin repeat homology 8e-26
 [SUPFAM] hypothetical protein MJ1322 4e-13
 [SUPFAM] kinesin-related protein KIF1B 3e-28
 [SUPFAM] kinesin motor domain homology 2e-63
 [SUPFAM] kinesin-related protein KLPA 7e-25
 [SUPFAM] kinesin-related protein nodA 1e-12
 [SUPFAM] kinesin-related protein Eg5 5e-30
 [PROSITE] ATP_GTP_A 1
 [PFAM] Kinesin motor domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW COMPLEXITY 7.53 %
 [KW] COILED_COIL 19.78 %

SEQ MESNFNQGVPRPSYVFSADPIARPSEINFDGIKLDLSHEFSLVAPNTEANSFESKDYLO
 SEG
 COILS
 3kar-
 SEQ VCLRIRPFTQSEKELESEGCVHILDSQTVVLKEPQCILGRLEKSSGQMAQKFSFSKVFG
 SEG
 COILS
 3kar-
 SEQ PATTQKEFFQGCIMQPVKDLLKGQSRIFTYGLTNSGKTYTFQGTEENIGILPRTLNVLF
 SEG
 COILS
 3kar-
 SEQ DSLQERLYTKMNLKPHRSREYLRSSSEKEEIASKSALLRQIKEVTVHNDSDDTLYGSL
 SEG
 COILS
 3kar-
 SEQ TNSLNISEFEESIKDYEQANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKML
 SEG
 COILS
 3kar-EEEEEEEEETTEEEETTCC-----CCEE
 SEQ RLSQDVKGYSFIKDLQWIVQSDSKEAYRLKLGIKHQSVAFKLNASSRSHSIFTVKIL
 SEG
 COILS
 3kar- EEETTTTTE-EEETTCCEEECCGGGHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEE
 SEQ QIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGERLRETGNINTSLTLGKCINVLKNS
 SEG
 COILS
 3kar- E--EETTTTCEEEEEEEEEEECCCCC---CCCHHHHHHHHHHHHHHHHHHHHHHTT
 SEQ EKSFKQHVFPFRESKLTHYFQSFNGKGKICMIVNISQCYLAYDETNLVLFSAIAQKVC
 SEG
 COILS
 3kar- TTTT--TCCTTTTTHHHHHHGGGCTTTTEEEEEEECCGGGHHHHHHHHHHHH.....
 SEQ VPDTLNSSQDKLFGPVKSSQDVSLDSNSNKKILNVKRATISWENSLEDLMEDEDLVEEIE

```

SEQ      .....XXXXXXXXXXXXXXXXXXXX
COILS
3kar-

SEQ      NAEETQNVETKLLDEDLKTLLENKAFISHEEKRKLLDLIEDLKKKLINERKEKLTLEFK
SEG      .....XXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXXXXXXXXXXXXXXXXX
COILS      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      IREEVTQEFYQYWAQREADFKETLLQEREILEENAERRLAIFKDLVGKCDTREAAKDIC
SEG      .....
COILS      CCCCCC
3kar-

SEQ      ATKVTEEATACLELKFNQIKAEALAKTKGELIKTKEELKKRENESDSLQIELETSNKKI
SEG      .....
COILS      .....CCCCCCCCCCCCCCCC
3kar-

SEQ      TQNQRIKELINIIQKEDTINEFQNLKSHMENTFKCNDKADTSSLIINNKLICHNETVEVP
SEG      .....
COILS      CCCCCCCCCCCCCC
3kar-

SEQ      KDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKEEVRPNIAEIEDIRVL
SEG      .....
COILS      .....CCCC
3kar-

SEQ      QENNEGLRAFLITIEENELKNEKEEKAELNKQIVHFQQELSLSEKKNLTLSEKVVQIQSNY
SEG      .....XXXXXXXXXXXXXXXXXXXX
COILS      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      DIAIAELHVQSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMHTKIDELRTLDSVSQ
SEG      .....
COILS      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      ISNIDLNLRLDSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAI
SEG      .....
COILS      .....
3kar-

SEQ      WEECKEIVKASSKSHQIEELEQQIEKLQAEVKGKYNENRLKEKEHKNQDDLKEKETL
SEG      .....XXXXXXXXXXXXXXXXXXXX
COILS      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      IQQLKEELQEKNVTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKVERSHS
SEG      .....
COILS      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      AKLEQDILEKESIILKLERNKEFQEHLDQSVKNTKDLNVKELKLEEITQLTNLQDMK
SEG      .....
COILS      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      HLLQLEEEETNRQETELKEELSASSARTQNLKADLQKREEDYADLKEKLTDAKKQIK
SEG      .....XXXXXXXXXXXXXXXXXXXX
COILS      CCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      QVQKEVSMRDEDKLLRIKINELEKKKNQCSQELDMQRTIQQLKEQLNNQVVEAIIQY
SEG      .....
COILS      CCCCCCCCCC
3kar-

SEQ      ERACKDLNVKEKIIEDMRMTLEEQEQTOVEQDQVLEAKLEEVRATELEKWKKECNLE
SEG      .....XXXXXXXXXXXXXXXXXXXX
COILS      .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ      TKNNQRSNKEHENNTDVLGKLTNLQDELQEQSKYNADRKKWLEEKMLLITQAKEAENIR
SEG      .....
COILS      CC
3kar-

SEQ      NKEMKKYAE DRERFFKQONEMEILTAQLTEKDSLQKWEERDQVAALEIQLKALISSN
SEG

```

```

COILS .....
3kar- .....

SEQ VQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGS
SEG .....
COILS .....
3kar- .....

SEQ VVLDSCSEVSTENDQSTRFPKPELEIQFTFLQPNKMAVKHPGCTTPVTVEIPKARKRSNE
SEG .....
COILS .....
3kar- .....

SEQ MEEDLVKCNKKKNATPRTNLKFPISSDRNSSVKKEQKVAIRPSSKITYSLRSQASIIGVN
SEG .....
COILS .....
3kar- .....

SEQ LATKKKEGTLQKFGDFLOHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKL
SEG .....
COILS .....
3kar- .....

SEQ YTSEISSPIDISGVILMDQMKESDHQIKRRLRTKTAK
SEG .....
COILS .....
3kar- .....

```

Prosite for DKFZphtes3_35b4.3

PS00017 152->160 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_35b4.3

```

HMM_NAME Kinesin motor domain
HMM *RCRPINeREindgcscvVQWpPwtGyktvnhghegds.....phks
R+RP+ + E++ + +V + +++++ ++ + ++
Query 64 RIRPFTQSEKELESEGCVHILDSQTIVLKEPQCILGRLEKSSGQMAQK 112
HMM FcFDHVFVWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQTGSGKTYTM
F+F +VF++++TQ++ +++ + V+D+++G IF+YG T SGKTYT
Query 113 FFSKVFVGPATTQKEFFQGCIMQPVKDLLKGQSRIFTYGLTNSGKTYTF 162
HMM MGpggehPDHmGIIPrcCHDIFdrIdkfgekDhdFW.....
G +++GI+PR+++ +ED++ + +++
Query 163 QG---TEENIGILPRTLNVLFDSLQERL-YTKMNLKPHRSREYLRSLSE 207
HMM .....
Query 208 QEKEEIASKSALLRQIKEVTVHNDSDDTLYGSLTNSLNISEFEESIKDYE 257
HMM .....hvkCSYMEIYNeeIYDLLCPnP...qhMkpLnIHEHPN
+V +S++EIYNE+IYDL +P++ Q++K L++ + +
Query 258 QANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQRRKMLRLSQDVK 307
HMM MGpYVqGCTEfHVCsYeDachWIWqGnknRHVAaTnMndhSSRSHtIFTI
+++++ V +A +++ +G K+ VA T++N SSRSH+IFT+
Query 308 GYSFIKDLQWIQVSDSKEAYRLLLKLGIKHQSVAFTKLNNASSRSHSIFTV 357
HMM HVeQrHk.qcdehvcHSKMNLVDLAGSERvniTGAEGQRlKEGcNINqSL
++ Q + + +++S ++L DLAGSER+ +T+ EG RL+E +NIN SL
Query 358 KILQIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGERLRETGNINTSL 407
HMM ttLGnVInaLaDgqTKYmYgghgHIPYRDSKLTWlLQDSLGGNcKtCMIA
+TLG++IN+L + + + +H+P+R+SKLT+ +Q + G +K CMI+
Query 408 LTLGKCINVLKNSE---KSKFQQHVPFRESKLTHYFQSFFNGRGKICMIV 454
HMM CIWPadWNYEETLSTLRYAdRAKnIkNkPQINEDPca*
+I+ + Y+ETL++L++ + A+++ + ++N+++++
Query 455 NISQCYLAYDETlnVLKFSaIAQKVCVPDTLNSSQDK 491

```

DKF2phtes3_35b5

group: metabolism

DKF2phtes3_35b5 encodes a novel 466 amino acid protein, with similarity to bovine accessory subunit for vacuolar ATPase and rat C7-1 protein.

The vacuolar proton-ATPase (V-ATPase) translocates protons into intracellular organelles or across the plasma membrane of specialized cells. The catalytic domain consists of a hexamer of 3 A subunits and 3 B subunits, plus accessory subunits C, D, and E. The rat homolog C7-1 seems to be enriched in aged adult rats in the frontal cortex.

The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A

complete cDNA, complete cds potential start at Bp 8, EST hits matches perfect to I54197 hypothetical protein, but possess 186 aa additional at N-terminus

Sequenced by DKFZ

Locus: unknown

Insert length: 2043 bp

Poly A stretch at pos. 2033, polyadenylation signal at pos. 2012

```
1  GCGGCCCATG GCGACGGCTC GAGTGGCGAT GGGGCCGCGG TCGCCCCAGG
51  CGCTCTGGCG CATGCCGTGG CTGCCGGTGT TTTGTCTGTT GCGGCCGCGG
101 GCGGCCGCGG CAGCGCGGGA GCAGCAGGTC CCGCTGGTGC TGTGGTCGAG
151 TGACCGGGAC TTGTGGGCTC CTGCCGCCGA CACTCATGAA GGCCACATCA
201 CCAGCGACTT GCAGCTCTCT ACCTACTTAG ATCCCGCCCT GGAGCTGGGT
251 CCCAGGAATG TGCTGCTGTT CCTGCAGGAC AAGCTGAGCA TTGAGGATTT
301 CACAGCATAT GCGCGTGTGT TTGGAACAA GCAGGACAGC GCCTTTTCTA
351 ACCTAGAGAA TGCCCTGGAC CTGGCCCCCT CCTACTGGT GCTTCTGCCC
401 GTCGACTGGT ATGCAGTCAG CACTCTGACC ACTTACCTGC AGGAGAAAGCT
451 CGGGGCCAGC CCCTTGCACTG TGACCTGGCC CACCCTGCGG GAGCTGAAGC
501 TCAATGCCAG CCTCCCTGCT CTGCTGCTCA TTGCGCTGCC CTACACAGCC
551 AGCTCTGGTC TGATGGCACC CAGGGAAGTC CTCACAGGCA ACGATGAGGT
601 CATCGGGCAG GTCCCTGAGCA CACTCAAGTC CGAAGATGTC CCATACACAG
651 CGGCCCTCAC AGCGGTCCGC CCTTCCAGGG TGCCCCGTGA TGTAGCCGTG
701 GTGGCCGGAG GCGTAGGTCT CCAGCTGCTA CAAAAACAGC CAGTATCACC
751 TGTGATCCAT CCTCCTGTGA GTTACAATGA CACCGCTCCC CGGATCCTGT
801 TCTGGGCCCA AAACCTTCTT GTGGCGTACA AGGACCAGTG GGAGGACCTG
851 ACTCCCTCA CCTTTGGGGT GCAGGAATCT AACCTGACTG GCTCCTCTGT
901 GAATGACTCC TTTGCCAGGC TCTCACTGAC CTATGAACGA CTCCTTGGTA
951 CCACAGTGAC ATTCAAGTTC ATTCTGGCCA ACCGCTCTA CCCAGTGTCT
1001 GCGCCGCACT GGTTTACCAT GGAGCCCTCT GAAGTCCACA GCAATGGCTC
1051 CGTCGGCTAC TTCAATGCTT CCGAGGTCAC AGGGCCAGC ATCTACTCCT
1101 TCCACTGCGA GTATGTCAGC AGCCTGAGCA AGAAGGGTAG TCTCTCTGTG
1151 GCGCCGACGC AGCCCTCTCC CTGGCAGATG ATGCTTCAGG ACTTCCAGAT
1201 CCAGGCTTTC AACGTAATGG GGGAGCAGTT CTCCTACGCC AGCGACTGTG
1251 CCAGCTTCTT CTCCCCGGGC ATCTGGATGG GGCTGCTCAC CTCCTGTGTC
1301 ATGCTCTTCA TCTTCACCTA TGGCCTGCAC ATGATCTCTA GCCTCAAGAC
1351 CATGGATCCG TTTGATGACC ACAAGGGCCC CACTATTCTT TTGACCCAGA
1401 TTGTGTGACC CTGTGCCAGT GGGGGGGTTG AGGGTGGGAC GGTGTCCGTG
1451 TTGTGTGCTT CCCACCTGTC AGCGCACTGG ACTGAAGAGC TTCCTCTTTC
1501 CTACTGCAGC ATGAACCTGA AGCTCCCTCT AGCCATCTTT GCTCCCTCTT
1551 CAGCCCGCTG AGGAGCTTTC TTGGGCTGCC CCCATCTCTC CCACCAAGGT
1601 GTACATATTC TGCGTAGATG CTAGACCAAC CAGCTTCCCA GGGTTCGTCT
1651 CTGTGAGGCG TAAGGGACAT GAATTTAGG GTCTCCTTTC TCCTTATTTA
1701 TTCTTGTGGC TACATCATCC CTGGCTGTGG ATAGTGTCTT TGTGTAGCAA
1751 ATGCTCCCTC CTTAAGGTAA TAGGGCTCCC TGAGTTGGG AGTGTGGAAG
1801 TACTACTTAA CTGTCTGTCC TGCTTGGCTG CCCTTATCTT TTTCTGGTGA
1851 TGTGTGTGTA ACAATAAGAA GTACACGGGT TTATTTCTGT GGCCTGAGAA
1901 GGAAGGGACC TCCACGACAG GTGGGCTGGG TGCGATCCCC GGCTGTTTGG
1951 CATGTTTCCA CCGGAGGTGC CGGGCAGGAG CATGGGGTGC TTGGTTGTTT
2001 CCTTCCTAAT AAAATAAACG CGGGTCGCCA TGCAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

95014142:
A novel accessory subunit for vacuolar H(+)-ATPase from chromaffin granules.

97215246:
Identification of a rat brain gene associated with aging by PCR differential display method.

Peptide information for frame 2

ORF from 8 bp to 1405 bp; peptide length: 466
Category: strong similarity to known protein

```

1 MATARVRMGP RCAQALWRMP WLPVFLSLAA AAAAAAAEQQ VPLVLWSSDR
51 DLWAPAADTH EGHITSDQLQ STYLDPALEL GPRNVLLFLQ DKLSIEDFTA
101 YGGVFGNKKQ SAFSNLENAL DIAPSSLVLP AVDWYAVSTL TTYLQEKLGA
151 SPLHVDLATL RELKLNASLP ALLLIRLPYT ASSGLMAPRE VLTGNDEVIG
201 QVLSTLKSED VPYTAALTAV RPSRVARDVA VVAGGLGRQL LQKQPVSPVI
251 HPPVSYNDTA PRILFWAQN FSVAYKDQWED LPTLFCGVQ LNLTSFVNDSFA
301 SFARLSLTIE RLFGTTFVK FILANRLYPV SARHWFMTMR LEVHNSGVSVA
351 YFNASQVTGP SIYSFHCEYV SLSKKGSL VARTQPSWQ MMLQDFQIQ
401 FNVMGQFSY ASDCASFFSP GIMMGLLTSL FHLFIPTYGL HMILSLKTHD
451 RFDDHKGPTI SLTQIV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b5, frame 2

TREMBL:AF035387.1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds., N = 1, Score = 2088, P = 3.8e-216

PIR:A55116 vacuolar ATPase (EC 3.6.1.-) chain Ac45 - bovine, N = 1, Score = 2011, P = 5.5e-208

PIR:I54197 hypothetical protein - human, N = 1, Score = 1464, P = 5.1e-150

>TREMBL:AF035387.1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds.
length = 463

HSPs:

Score = 2088 (313.3 bits), Expect = 3.8e-216, P = 3.8e-216
Identities = 408/463 (88%), Positives = 426/463 (92%)

```

Query:      4 ARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAEQQVPLVLWSSDRDLWAPAADTHEGH 63
            +R+R G R A LW + LSL A AAA AAEQQVPLVLWSSDRDLWAP ADTHEGH
Sbjct:      8 SRIRTGTRWAPVLW-----LLSLVAVAAVAAEQVPLVLWSSDRDLWAPVADTHEGH 61

Query:     64 ITSDQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 123
            ITSD+QLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA
Sbjct:     62 ITSDMQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 121

Query:     124 PSSLVLPVADWYAVSTLTITTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 183
            PSSLVLPVADWYA+STLTITTYLQEKLGASPLHVDLATL+ELKLNASLPALLLIRLPYTASS
Sbjct:     122 PSSLVLPVADWYAISTLTITTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 181

Query:     184 GLMAPREVLTGNDEVIGQVLSTLSEDEVPTAALTAVRPSRVARDVAVVAGGLGRQLLQK 243
            GLMAPREVLTGNDEVIGQVLSTL+SEDEVPTAALTAVRPSRVARDVA+VAGGLGRQLLQ
Sbjct:     182 GLMAPREVLTGNDEVIGQVLSTLSEDEVPTAALTAVRPSRVARDVAVVAGGLGRQLLQT 241

Query:     244 QPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKDQWEDLPTLFCGVQELNLTGSFWNDSFA 303
            Q SP IHPPVSYNDTAPRILFWAQNFSVAYKD+W+DLT LTFGV+ LNLTSFVNDSFA
Sbjct:     242 QVASPAIHPPVSYNDTAPRILFWAQNFSVAYKDEKDLTSLTFGVENLNLTSFVNDSFA 301

Query:     304 RLSLTIERLFGTTVTFKIFILANRLYPVSARHWFMTMERLE+HSNGSVA+FN SQVTGPSIY 363
            LSLTYE LFG TVTFKIFILA+R YPVSAR+WFMTMERLE+HSNGSVA+FN SQVTGPSIY

```

Sbjct: 302 MSLTYEPLFGATVTFKFIASRFYPSARYWFTMERLEIHSNGSVAHFNVSQVTGPSIY 361
Query: 364 SFHCEYVSSLSKKGSLLVARTQSPWQMMLQDFQIQAFNVMEQFSYASDCASFFSPGIW 423
SFHCEYVSSLSKKGSLLV PS WQM L +FQIQAFNV GEQFSYASDCA FFSPGIW
Sbjct: 362 SFHCEYVSSLSKKGSLLVNTV-PSLWQMTLHNFQIQAFNVTEQFSYASDCAGFFSPGIW 420
Query: 424 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV 466
MGLLT+LFLMFIFTYGLHMILSLKTMDFDD KGPTI+LTQIV
Sbjct: 421 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTITLTQIV 463

Pedant information for DKF2phtes3_35b5, frame 2

Report for DKF2phtes3_35b5.2

[LENGTH] 466
[MW] 51621.44
[pI] 5.73
[HOMOL] TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1
protein (C7-1) mRNA, complete cds. 0.0
[PIRKW] hydrolase 0.0
[PROSITE] MYRISTYL 7
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 7
[KW] SIGNAL PEPTIDE 38
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 11.59 %

SEQ MATARVMGPRCAQALWRMPWLPVFLSLAAAAAAAEQVPLVLWSSDRDLWAPAADTH
SEGXXXXXXXXX
PRD cccccccccchhhhhccccchhhhhhhhhhhhhhhhhcccccccccccccccccc
MEM
SEQ EGHITSDLQLSTYLDPALELGPRLVLLFQDKLSIEDFTAYGGVFGNKQDSAFSNNLENAL
SEGXXXXXXXXXXXXXXXXXXXX
PRD cccccchhhhhccccccccccccccccccccccccccccccccccccchhhhhhhcc
MEM
SEQ DLAPSSLVLPADVWYAVSTLTLYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYT
SEGXXXXXXXXXXXXXXXXXXXX
PRD cccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccc
MEM
SEQ ASSGLMAPREVLTGNDEVIGQVSTLKSSEDVPTYAALTAVRPSRVARDVAVVAGGLGRQL
SEGXXXXXXXXXXXXXXXXXXXX
PRD cccccccccccccccccchhhhhccccccccchhhhhhhccccccccccccccccchh
MEM
SEQ LQKQPVSPVHPVSYNDTAPRILFWAQNFSVAYKDQWEDLPLTFGVQELNLTGSFWND
SEG
PRD hhhhhcc
MEM
SEQ SFARLSLTYERLFGTTVTFKFIANRLYPVSARHWFTMERLEVHSNGSVAYFNASQVTGP
SEG
PRD hhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcccccccccccccccccc
MEM
SEQ SIYSFHCEYVSSLSKKGSLLVARTQSPWQMMLQDFQIQAFNVMEQFSYASDCASFFSP
SEGXXXXXXXXX
PRD cccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccc
MEMMMMMM
SEQ GIWMGLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV
SEG
PRD ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc
MEM MMM

Prosites for DKF2phtes3_35b5.2

PS00001	166->170	ASN_GLYCOSYLATION	PDOC00001
PS00001	257->261	ASN_GLYCOSYLATION	PDOC00001
PS00001	269->273	ASN_GLYCOSYLATION	PDOC00001

WO 01/12659

PCT/IB00/01496

PS00001	292->296	ASN_GLYCOSYLATION	PDOC00001
PS00001	299->303	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00004	375->379	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	331->334	PKC_PHOSPHO_SITE	PDOC00005
PS00005	374->377	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	114->118	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00007	207->214	TYR_PHOSPHO_SITE	PDOC00007
PS00008	102->108	MYRISTYL	PDOC00008
PS00008	103->109	MYRISTYL	PDOC00008
PS00008	200->206	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35b5.2)

DKF2phtes3_35e21

group: differentiation/development

DKF2phtes3_35e21.2 encodes a novel 104 amino acid putative interleukin precursor, related to interleukin-7.

Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells.

This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

similarity to interleukin-7 precursor

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2095 bp

Poly A stretch at pos. 2085, polyadenylation signal at pos. 2067

```
1 GGATGAAAGT GATTAAATC ATTTTAGAA TTTTTTTT GTTTGTTTT
51 AGCAACATGC TGAACAATA ATTTACTTA AAAATAAGCC AGTTAAAAA
101 AAGGACGCTA AGCCCAAGTG GGGGGCAATA TTAGTCAGGA TCCTTGGGGT
151 CTAATTCAG ACCAACTTC AGAAGCACTT CTTGTCTCT GTTCTCACCT
201 CTGCTGCCCT TCCTTCCCT CATCCCCTAA GAGAGACAAA GATAAAGGCC
251 CACCTGCATC CCTAAGTCT ACTGAGATCA GCCACCCAG GGGAGAGAAA
301 CTGGATCTAC TTACAGCCAC CCCCTGTTT CATCCATATA CTTACTTCCC
351 CCAATTTGCA TGTGATTATG GAAACAAGTC ATGCTCATGA AAGCAACTGT
401 AAAATAAAG GTTATGGAGT AGTTACAGAA CTTCTTCACA GCCAGCTTGG
451 TGGAGCTGGG GAGGACTTAG GGGCCATTGG AGTCTCTTAT GTGTACAGCT
501 TCAGGCGTGT CCCTTTCAGT TTGATTTTAA GCAATGCCTC ACTTCATAGC
551 TTAGGGGGTA AGGATTCCAT TCAGGTAGGT TGTCTAAAGG AACTAATGGG
601 ACCTCTCAGT GAATTAGCTG ACCAGATTTT AGGAAATCTT TTTAATTTCT
651 ATGATTTTCC TTCTCACATT TTGAAATGGT AAAATTGACT GGAAATAATT
701 TTCTTGGTG CCTTATTGGT TTCTCTTGCA AACCTTTCTC ATATTTTCTC
751 ATGACCATTG CCAGTGACCA AGGCCCATGT GTGTGTTGTG TGTAAATGTG
801 GGCATGTACA AGCTTAAATA ACSTGCCGAC AGCAGCTGTT CAAAGTTGGT
851 ATTCATTAGG CTGTTGCCCT CTGGGCTGGA GCTGGCTTAA TCCTGACACC
901 GGCTGCCAGG AGAAAACCTC ATGGATCACA CACCAACCTT TAATAACAGC
951 ATCCGTGACC TGCACCTCC AGTACAGAAT CCCAGCCAAA GAAGCTTCTT
1001 ATGTAGTTGT ATATTTTAAAT GAATCCTTAC AAGCTTCTCT GAGCTTCTT
1051 TCACCTTTGT GCTCTACAGA AAGCCCAAGG GGGTAGGAG GCACAGAGCT
1101 TTGAATAACT CTTTCTAAC ACTAAATCTG CCAACAGGA CAGAGCACAT
1151 CACACGTATA GGCAGGTGTG AGGACAGTG GCTAAGAATT GCCTGCTCCC
1201 TCTCATGCT CTTTCTGTT TCCTAAGTCC AATCAAGTGA TCCTGGGAAA
1251 CAAATCTGTC TGGATTGCGG AGGGTGGTTC TGAAGAAGCT GCCAAGACGT
1301 TAAAGAAGGG TGAAGAGTAG GCAGAATAA AGTAGTAAC CTGAGTCAAG
1351 ACTCTCAAAA GCTAGCAGCC TGATGACAAT AGGATTTATT TCAGCCAGGA
1401 TAGTGTCTGT CTGTGAGTGC ATCATTTTAA GACAGTATGA CTTCATGTTG
1451 TTACAAACTA TGTATAGTAT GTATGTTTGG TGGGTGTGAT ATATACATAA
1501 TATATATTAT ATATATATAT GAGAGATTG GTGACTTTTG ATACGGGTTT
1551 GGTGCAGGTG AATTTATTAC TGAGCCAAAT GAGGCACATA CCGAGTCAGT
1601 AGTTGAAGTC CAGGGCATTG GATACGTGTT ATGATTTCCT TATATGTATA
1651 GTGCCTATCC CATGCTGTAG TCACTGTTAT GTTAAATCCA GAAGTTACAC
1701 TAGAGCCAGC GATACCTTAT TTGTAGACAA TCAATTTGAA TCCATATGTT
1751 ATTACTGGCA GATGATACAT GATTACAGT CTGAATCTGT AACACTTACA
1801 AAAGGAAACC CAGAGCAGCT TGATGAGTTT TTGTTCTGCT TTCGTTCTCG
1851 GGAGTCAGTA GAAACAGCAG TTGTATGTGG TTATGTTAGT CTCAAGATAC
1901 TTAATTTGTT GACCTTACTT CAGAAAAATT TTGTATGTAT TATATTTGTG
1951 GGAAGGTAAA ATAATCATT GAGATTTTTA TCAATATGA AGATTAGTTA
2001 TTTATGAAAA ACAAGAAAT GTCTATTTT CTTGTTCCTC AATTAATGTA
2051 GATAAATTTT AAAATGCATT AAGTAATGCG TCCGGAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

89098903:
Human interleukin 7: molecular cloning and growth factor
activity on human and murine B-lineage cells.

Peptide information for frame 2

ORF from 368 bp to 679 bp; peptide length: 104
Category: similarity to known protein

1 METSHAHESN CKIKGYGVVQ QLLHSQLCGA GEDLGPVGVS YVYSFRAVPF
51 SLILSNASLH SLGGKDSIQV GCLKELMGPL SELADQILGN LFNFDFFSH
101 ILKW

BLASTP hits

Entry B32223 from database PIR:
interleukin-7 precursor (clone 1) - human
Score = 66, P = 7.0e-01, identities = 21/70, positives = 33/70

Alert BLASTP hits for DKFZphtes3_35e21, frame 2

PIR:B32223 interleukin-7 precursor (clone 1) - human, N = 1, Score =
66, P = 0.72

TREMBL:PADAL1_1 gene: "dall"; P.abies dall mRNA, N = 2, Score = 59, P
= 0.77

PIR:C32223 interleukin-7 precursor (clone 4) - human, N = 1, Score =
66, P = 0.79

TREMBL:PRU76726_1 gene: "PrMADS3"; product: "MADS-box protein"; Pinus
radiata MADS-box protein (PrMADS3) mRNA, complete cds., N = 2, Score =
59, P = 0.94

>PIR:B32223 interleukin-7 precursor (clone 1) - human
Length = 133

HSPs:

Score = 66 (9.9 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 21/68 (30%), Positives = 33/68 (48%)

Query: 39 VSYVYSFRAVPFSLIL-----SNASLHSLGGK--DSIQVGCLKELMGPLSELADQILGNL 91
VS+ Y F P L+L S+ + GK +S+ + +L+ +E+ L N
Sbjct: 4 VSFRIYFGLPPLILVLLPVASSDCDIEGKDGKQYESVLMVSIQQLLDSMKIEGNSCLNNE 63

Query: 92 FNFYDFPSHI 101
FNF F HI
Sbjct: 64 FNF--FKRHI 71

Pedant information for DKFZphtes3_35e21, frame 2

Report for DKFZphtes3_35e21.2

[LENGTH]	104
[MW]	11339.12
[pI]	5.87
[PROSITE]	MYRISTYL 2
[PROSITE]	PKC_PHOSPHO_SITE 1
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	Alpha_Beta

SEQ METSHAHESNCKIKGYGVVQQLLHSQLCGAGEDLGPVGVS YVYSFRAVPFSLILSNASLH
PRD ccchhhhhcccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccc

WO 01/12659

PCT/IB00/01496

SEQ SLGGKDSIQVGLKELMGPLSELADQILGNLFNFYDFPSHILKW
PRD cccccceccccccccchhhhhhhcccccccccccccc

Prosite for DKFZphtes3_35e21.2

PS00001	56->60	ASN_GLYCOSYLATION	PDOC00001
PS00005	44->47	PKC_PHOSPHO_SITE	PDOC00005
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35e21.2)

DKFZphtes3_35g6

group: testes derived

DKFZphtes3_35g6 encodes a novel 482 amino acid protein with high partial similarity to H. sapiens chromosome 19, cosmid R27216.

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to R27216_1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="15"

Insert length: 3177 bp

Poly A stretch at pos. 3167, polyadenylation signal at pos. 3148

```
1 GGAGGCAGCG CCGGCCTCCG GAGGCGGCCT GGGCGATGGC GCGGAGTTT
51 TGTCCATAAC CTGGGCAACC GCGCAGCTGG AGGATGGCCT CACTCGGGCC
101 TGCCGCGAGCT GGGGAGCAGG CGTCGGGGGC TGAGGCGGAG CCGGGCCCCG
151 CGGGGCCGCC GCGCCGCCCC TCACCGTCTT CTCTGGGGCC CCTGCTCCCC
201 CTGACGCGGG AACCTCTCTA CAACTGGCAG GCGACCAAGG CGTCGCTGAA
251 GGAGCGCTTC GCCTTCCTCT TCAACTCGGA GCTGCTGAGC GATGTCGCT
301 TCGTACTGGG CAAGGGTCGC GCGCGCGCCG CCGCTGGGGG CCGCAGCGCC
351 ATCCCCGCCC ACCGCTTCGT GCTGGCGGCC GGCAGCGCCG TCTTTGACGC
401 CATGTTCAAC GCGGCGATGG CCACCACGTC GGCAGAGATC GAGCTGCCGG
451 ACGTGGAGCC CGCAGCCTTC CTGGCGCTGC TGAGATTTCT ATATTGAGAT
501 GAAAGTTCAAA TTGGTCCAGA AACAGTTATG ACCACTCTTT ATACTGCCAA
551 GAAATACGCA GTCCAGCCCT TGGAGGCACA CTGTGTAGAA TTCTCCACCA
601 AACATCTTAG GGCAGATAAT GCCTTTATGT TACTTACTCA GGCTCGATTA
651 TTTGATGAAC CTCAGCTTGC TAGTCTTTGT CTAGATACAA TAGACAAAAG
701 CACAATGGAT GCAATAAGTG CAGAAGGGTT TACTGATATT GATATAGATA
751 CACTCTGTGC AGTTTATAGG AGAGACACAC TCAGTATTCG AGAAAGTCGA
801 CTTTTTGGAG CTGTTGTACG CTGGGCAGAA CGAGAATGTC AGAGACAACA
851 ATTACCTGTG ACTTTTGGGA ATAAACAAAA AGTTCTAGGA AAGCACTTT
901 CCTTAATCCG GTTCCCACTG ATGACAAATT AGGAATTTGC AGCAGGTCTT
951 GCTCAATCTG GAATTTGTCT AGATCGTGAA CTGGTAAACC TCTTCTCTCA
1001 TTTTACTGTG AACCTTAAC CCAGAGTTGA ATACATTGAC CGACCAAGAT
1051 GCTGTCTCAG GGGAAAGGAA TCGTCATCA ATAGATTCCA CCAAGTAGAA
1101 AGCCCGTGGC GTTACAGTGG GACGAGTGAT CGAATCAGAT TCACAGTTAA
1151 TAGAAGGATC TCTATAGTTG GATTGGCTTT GTATGGATCT ATTCATGGCC
1201 CTACAGATTA TCAAGTGAAT ATACAGATCA TTGAATATGA GAAAGAGCAA
1251 ACCCTGGGAC AGAATGATAC CGGCTTTAGT TGTGATGGGA CAGCTAACAC
1301 ATTCAGGCTC ATGTTCAAGG AACCCATAGA GATCCTGCCC AATGTGTGCT
1351 ACACAGCATG TGCAACACTC AAAGGTCCAG ATTCCCACTA TGGCACAATA
1401 GGATTGAAGA AAGTAGTGCA TGAGACACCT GCTGCAAGCA AGACTGTTTT
1451 TTTCTTTTTT AGTTCCCTCG GCAATAATAA TGGCACTTCA ATAGAAGATG
1501 GACAAATCCC AGAAATCATA TTTTATACAT AATTTAGCAT TATAATACAT
1551 CTTGGCTAAA TAATACCATA CAATCTAGTG TCAAAAACAT AAATGGCCAC
1601 AAAAAAGTAG TTTGAGTGTT ATGAATATTT AAAATTGTAA GATAAGAAAC
1651 AGTTTCTTAG AGCAGATAGA AAAATGCTTA TTTAAATCTT TGCAATGATT
1701 AAAAAACAGT TTTCCATTTT CTTACAACCT TAAGAGAAAA GAACTGGGTT
1751 TAATGGTTTA AAAAAAGCA CAGCTTTTTC ACCTTCATCT TGTATAATTT
1801 CATAGATTGG CTGACTAGG GTCTTCAAT AGTTTGGGAA TTGAAAGATT
1851 CTTGTTATAT ATAGCTAGTT TGGGTTTGT TTTGTTTAA CTATTTTGAA
1901 GGTTAGTGGA GATGGGCAAA TAGGCTTAAC TATTTTGAA GTTGGATGAA
1951 AAGAGATGGG TCAGTATTC TACAGAATTC TTATTAACCT AAATAACTAA
2001 ATTCAGAAAA ATTAAGAAGC TGACTTTATA TTTGGTGGTT TGAAGTATCT
2051 TGTGTTAGC ATTTGTAATA ATGCTAAAAA AGGCCTAATA AAATGCCCAA
2101 GAAATATTC AGTGCAATTA TAGAAGAGGA TATTTGTAG TAGTATAGTA
2151 ATGTGTTATG TAGTACAGTT TTAAGCTAT AAATGGAATT TTGTGTAAT
2201 TCACAAAAAT GTGATATAAA CAGGATCTAA GACTGGATTC CCTGTCACTA
2251 AACTGCACCA CTATACCTGT CTCTCTGTGT GGGGGACACT GCTGATGATT
2301 CCCAAGATTG AGATGATGAC GGTGATGACG ACTGGGTGAA CAGCCATCAC
2351 TTCAACATTG TGATAATCCT TCACAGCAAG AAACCGAATA AAATACTAAC
2401 ATTTCTAACA ACTGCTCTGA CATTGTAAAG AGATCCAACA GAATCACTCC
2451 TGCTGAAAAA TACGCTTCTT GCCACCTACA CATTCTATT TAGGAAGTAA
2501 AATTTGCTTC ATGGTCATGA CCCCATTAGT CAGTGTATCA GCTGTGTTGG
2551 GGATAGGAAG TATATCTGGC AGATTGACAT TTATACACTT TTTTATAAAG
2601 CAGATTTTAA AATATAGTAA CATCCATTTT TTTCCCTTGA AAGTGATTCT
2651 CTTATAAAAA ATGAAAGTGG AGTTTAAAGT ATATCAAAAT GTTGTTGGAAG
2701 GTGATTAAAA ATCAAAATTC TTTTAAATAT CAACTTAATT TTTTCTAAGT
```

```

2751 AAGATACAAA AAATTTTCAT CTAAAGTAAT ATTTCACTTT ATATTGTAAA
2801 GAAGGTAGGT ATATTGGTGG CTGAGGTCTC TTGAAATTGC TAAAGGGAAA
2851 TTTTCTATG GTAATGCTCT TACGGATATA AGCCTCAGTT AAATGGAAAT
2901 ATCTATGGGA TGTGTGGTTC TGGTTAACTA AAAATTAACC AGTAAACACT
2951 CTGTAGTAAC CATTACAGAA AATACTTCTG CCTTAAAAAA TATGATATGC
3001 CAGAGATGAG TTAGTGTTC TTGACGTGG AGACCTATAA ATGCCTCATC
3051 TGTGTACTG AACAAATTGAA ACTGCATGCA GCCATAAAG GGACAAGAAA
3101 CAGAACTGTT TACTAACTTT GGGACATCCC CTGGAGTTTT TAAAAATAAA
3151 TAAATATATA TATATATAAA AAAAAA

```

BLAST Results

Entry G37753 from database EMBL:
 SHGC-63477 Human Homo sapiens STS genomic.
 Score = 1627, P = 3.0e-66, identities = 327/329

Entry G37752 from database EMBL:
 SHGC-63476 Human Homo sapiens STS genomic.
 Score = 1578, P = 6.2e-64, identities = 320/324

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 84 bp to 1529 bp; peptide length: 482
 Category: similarity to unknown protein

```

1 MASLGPAAG EQASGAEEF GPAGPPPPPS PSSLGPLLPL QREPLYNWQA
51 TKASLKERFA FLFNSLELSD VRFVLGKGRG AAAAGGPQRI PAHRFVLAAG
101 SAVFDAMFNG GMATTSAEIE LPDVEPAFL ALLRFLYSDE VQIGPETVMT
151 TLYTAKKYAV PALEAHCVEF LTKHLRADNA FMLLTQARLF DEPQLASLCL
201 DTIDKSTMDA ISAEGFTDID IOTLCAVLER DTLISRESRL FGAVVRWAEA
251 ECQRQQLPVT FGNKQKVLGK ALSLIRFPLM TIEEFAAGPA QSGLSDREV
301 VNLFLHFTVN PKPRVEYIDR PRCLRGKEC CINRFQVES RWCYSGTSDR
351 IRFTVNRIS IVGFLYCSI HGPTDYQVNI QIIEVEKKQT LGQNDTGFSC
401 DGTANTFRVM FKEPIELPN VCYTACATLK GPDSHYGTRG LKKVVHETPA
451 ASKTVFFFFS SPGNNGTSTI EDGQIPEIIF YT

```

BLASTP hits

Entry AC005306.2 from database TREMBL:
 product: "R272I6.1"; Homo sapiens chromosome 19, cosmid R27216,
 complete sequence.
 Score = 1298, P = 1.9e-132, identities = 245/297, positives = 268/297

Entry CEF38H4.9 from database TREMBLNEW:
 gene: "F38H4.7"; Caenorhabditis elegans cosmid F38H4
 Score = 1237, P = 5.6e-126, identities = 248/446, positives = 322/446

Entry AC004678.1 from database TREMBL:
 product: "R34094.1"; Homo sapiens chromosome 19, cosmid R34094,
 complete sequence.
 Score = 555, P = 1.0e-53, identities = 112/137, positives = 123/137

Alert BLASTP hits for DKFZphtes3_35g6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35g6, frame 3

Report for DKFZphtes3_35g6.3

```

[LENGTH] 482
[MW] 52771.47
[pI] 5.79

```


[illegible]

PS000001	394->398	ASN GLYCOSYLATION	PD0C00001
PS000001	466->470	ASN GLYCOSYLATION	PD0C00001
PS000004	357->361	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	387->391	CAMP_PHOSPHO_SITE	PD0C00004
PS000005	54->57	PKC_PHOSPHO_SITE	PD0C00005
PS000005	154->157	PKC_PHOSPHO_SITE	PD0C00005
PS000005	234->237	PKC_PHOSPHO_SITE	PD0C00005
PS000005	296->299	PKC_PHOSPHO_SITE	PD0C00005
PS000005	348->351	PKC_PHOSPHO_SITE	PD0C00005
PS000005	406->409	PKC_PHOSPHO_SITE	PD0C00005
PS000006	428->431	CK2_PHOSPHO_SITE	PD0C00006
PS000006	14->18	CK2_PHOSPHO_SITE	PD0C00006
PS000006	54->58	CK2_PHOSPHO_SITE	PD0C00006
PS000006	115->119	CK2_PHOSPHO_SITE	PD0C00006
PS000006	206->210	CK2_PHOSPHO_SITE	PD0C00006
PS000006	217->221	CK2_PHOSPHO_SITE	PD0C00006
PS000006	234->238	CK2_PHOSPHO_SITE	PD0C00006
PS000006	281->285	CK2_PHOSPHO_SITE	PD0C00006
PS000006	296->300	CK2_PHOSPHO_SITE	PD0C00006
PS000006	468->472	CK2_PHOSPHO_SITE	PD0C00006
PS000007	439->447	CK2_PHOSPHO_SITE	PD0C00007
PS000008	80->86	MYRISTYL	PD0C00008
PS000008	110->116	MYRISTYL	PD0C00008
PS000008	365->371	MYRISTYL	PD0C00008

WO 01/12659

PCT/IB00/01496

PS00008	392->398	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	463->469	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35g6.3)

DKFZphtes3_35k16

group: metabolism

DKFZphtes3_35k16 encodes a novel 666 amino acid protein with weak similarity to fatty acid-CoA synthetaseses/ligases.

The novel protein contains a putative AMP-binding domain signature, which is present in enzymes, which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. Therefore it is a new fatty acid-CoA synthetaseses/ligase with unknown substrate.

The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

similarity to acyl-CoA synthetase

complete cDNA, complete cds, potential start codon at Bp 50, few EST hits, seems to be a testis specific cDNA, 5 of 6 EST hits are from testis derived libraries

Sequenced by DKFZ

Locus: unknown

Insert length: 2520 bp

Poly A stretch at pos. 2510, polyadenylation signal at pos. 2490

```
1 CAGATGTCCC AGCTCCAGTG CTGTGGAGCA TGGTTCTG ACACCTGGAA
51 TGACTGGAAC CCCAAGACT CAAGAAGGAG CTAAAGATCT TGAAGTAGAC
101 ATGAATAAAA CAGAAGTTAC TCCAGGCTG TGGACCACCT GTCGAGATGG
151 AGAAGTCCTT CTGAGGCTAT CCAACACGG ACCAGGCCAT GAGACCCCGA
201 TGACCATCCC TGAATTTTTT CGAGAGTCAG TCAACCGATT TGAACCTTAT
251 CCAGCCCTCG ATCCAGAGAA TGGCAAAAAG TGGGAATTC TGAATTTCAA
301 CCAGTACTAT CAGGCTCTC GGAAGCTCG AAAATCCTTG ATCAAGCTGG
351 GTTGGAGCGG TTTCCACGGA GTTGCTATCC TGGGGTTAA CTCTGCACAG
401 TGGTTATACA CTGCTGTGGG TGCCATCCTA GCCGGGGCTG TTTGTGTGG
451 TATTTATGCC ACCAACTCTG CCGAGGCTTG TCAATATGTC ATCACTCATG
501 CCAAGTGAA CATCTTGCTG GTTGAGATG ATCAACAGTT ACAGAAAATC
551 CTTTCGATTC CACAGAGCAG CTTAGAGCCC CTAAAAGCGA TCATCCAGTA
601 CAGACTGCCA ATGAAGAAGA ACAACAACCT GTACTCTTGG GATGATTTC
651 TGGAACTTGG CAGAAGTATC CCTGACACCC AACTGGAGCA GGTCTATCGAG
701 AGCCATAAGG CGAATCAATG CGCAGTGCTC ATCTACACTT CAGGGACCAC
751 AGGCATACCC AAGGGAGTGA TGCTCAGTCA TGACAACATC ACGTGGATTG
801 CAGGAGCAGT GACAAAGGAC TTAAACTGCA CAGACAAGCA TGAGACGGTG
851 GTTAGCTACC TCCCACTCAG CCATATTGCA GCACAGATGA TGGACATCTG
901 GGTACCCATA ARGATTGGGG CGCTCACATA CTTTGCTCAA GCAGATGCTC
951 TCAAGGGCAC CTGGTAAGT ACTCTAAAGG AGGTAAAACC TACTGTCTTC
1001 ATTGGAGTGC CTCAAATTTG GGAGAAGATA CATGAGATGG TGAAGAAAAA
1051 TAGTCCGAAG TCCATGGGCT TGAAGAGAA GGCATTCCTG TGGGCAAGAA
1101 ACATTGGGCTT CAAGGTCAAC TCAAAAAGA TGTGGGGAA ATATAATACT
1151 CCCGTGAGCT ACCGCATGGC TAAGACTCTC GTGTTCAGCA AAGTCAAGAC
1201 ATCCCTTGGC TTGGATCACT GTCACTCTTT TATCAGTGGG ACTGGCCCCC
1251 TCAACCAAGA GACTGCCGAG TTCTTTCTAA GCTTGGACAT ACCTATAGGC
1301 GAGTTGTATG GGTGAGTGA GAGCTCGGGA CCCCACACGA TATCCAACCA
1351 GAATAACTAC AGGCTTCTAA GCTGTGGCAA GATCTTGACT GGGTGAAGA
1401 ATATGCTGTT CCAGCAGAAC AAGGATGGCA TTGGGGAGAT CTGCCTCTGG
1451 GGTAGGCACA TCTTCATGGG CTATCTGGAA AGTGAGACTG AAACCTACAGA
1501 GGCCATCGAT GATGAAGGCT GGCTACACTC TGGGGATCTG GGCCAGCTGG
1551 ACGGTCTGGG TTTCTCTAT GTCAACGGCC ACATCAAAGA AATCCTTATC
1601 ACTGCTGGTG GTGAAAATGT GCCCCCCATT CCTGTTGAGA CCTTGTTAA
1651 GAAGAAGATC CCCATCATCA GTAACGCCAT GTTAGTAGGA GATAAACTGA
1701 AGTTTCTGAG CATGTTGCTG ACCTGAAGT GTGAGATGAA TCAGATGAGC
1751 GGAGAACCCT TGGACAAGCT GAACCTCGAG GCCATCAACT TCTGTGGGGG
1801 TCTGGGCAGC CAGGCATCCA CCCTGACTGA GATGGTGAAG CAGCAAGACC
1851 CCTGGTCTTA CAAGGCCATC CAGCAAGCCA TCAATGCTGT GACCAAGAAA
1901 GCCATGAACA ATGCACAGAG GATTGAAAAG TGGTCACTCT TGGAGAAGGA
1951 CTTTTCATC TATGCTGGAG AGCTAGGTCC AATGATGAAA CTTAGAGAC
2001 ATTTTGTAGC CCAGAAATAC AAAAAACAAA TTGATCACAT GTACCACTGA
2051 CTGCTTTGAT GGAGCTGCTC TCAGCTGTTT TGATGCCTTC AGCAGGAAGA
2101 CCTCATTGCA ATAAGTGAAA TGCTGCTCTA GGTAGAAGCT CTCCTGTCTG
2151 TTTTAAAGAA GCCACATTCC TCATTGGTCA GTTCTTGAT TGTTCGTCTG
2201 TTGGAGAGGT GCTCCCTAGA AGAACCTGCC ATACGTTTCA AAGCAATAAA
2251 ATCACTGTAT ATCTTTCTAA GGACCTTCAA GTCATGACTC CAGGGAAGCC
2301 TATTGGGAAG TCTACTAAAA ACTGCCTGAT TTACAAGAAA GACCTGAAC
```

2351 TGTGGGCTCC CATTGATTT TTTTCTCTC AGGGGACTCA GACATTAGAA
2401 AGAAAAAGCC TCACAGATTT GAAGAACTGG ACCCCCAAAAT CAACTACCT
2451 GCCTGGAAGC AACTGGGAAA CCCTTCCAAT AAGTCTGAT AATAAAGCAC
2501 TTCAGGGTCC AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2047 bp; peptide length: 666
Category: similarity to known protein

1 MTCTPKTQEG AKDLEVDNNK TEVTPRLWTT CRDGEVLLRL SKHGPGHETP
51 MTIPEFFRES VNRFGTYPAL ASKNGKKWEI LNFNOYYEAC RKAASLILK
101 GLERFHGVGI LGFNSAEWFI TAVGAILAGG LCVGIYATNS AEACQYVITH
151 AKVNILLVEN DQLOKILSI POSSLEPLKA IIQYRLPMKK NNNLYSWDDF
201 MELGRSIPDT OLEOVIESQK ANQCAVLIYT SGTGIPKGV MLSDNITWI
251 AGAVTKDFKL TDKHETVVS Y LPLSHIAAQM MDIWPVPIKIG ALTYFAQADA
301 LKGLTVSTLK EVKPTVFIGV PQIWEKIHQM VKKNSAKSMG LKKKAFVWAR
351 NIGFKVNSKK MLGKYNTFVS YRMAKTLVFS KVKTSGLGDH CHSFISGTAP
401 LNQETAFFEL SLDIPIGELY GLESSSGPHT ISNQNRYLL SCGKILTCK
451 NMLFQONKDG IGEICLWGRH IFNGYLESET ETTEAIDDEG WLHSGDLGQL
501 DGLGFLYVTG HIKEILITAG GENVPPIPVE TLVKKKIPII SNAMLVGDKL
551 KFLSMLTLK CEMNQMSGEP LDKLNFEAIN FCRGLGSQAS TVTEMVKKQD
601 PLVYKAIQGG INAVNQEAMN NAQRIEKWI LEKDFSIIYG ELGPMMLKLR
651 HFVAQRYKKQ IDHMYH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZpHtes3_35k16, frame 2

TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo
sapiens mRNA for KIAA0631 protein, partial cds., N = 1, Score = 1641, P
= 8.9e-169

PIR:E70937 probable fadD15 - Mycobacterium tuberculosis (strain H37RV),
N = 2, Score = 532, P = 3.6e-62

PIR:H64041 long-chain-fatty-acid--CoA ligase homolog - Haemophilus
influenzae (strain Rd KW20), N = 2, Score = 486, P = 6.5e-59

>TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo
sapiens mRNA for KIAA0631 protein, partial cds.
Length = 634

HSPs:

Score = 1641 (246.2 bits), Expect = 8.9e-169
Identities = 319/628 (50%), Positives = 440/628 (70%)

Query: 38 LRLSKHGPGHETPMTIPEFFRESVNRFGTYPALASKNGKKWEILNFNOYYEACRKAASL 97
LR+ P + P T+ F E++++G AL K RWE +++QY R+AAK
Sbjct: 2 LRIDPSCP--QLPYTVHRMFYEALDKYGDIALGFKRQDKWEHISYSQYLLARRAAGF 59

Query: 98 IKLGLERFHGVGILGFNSAEWFI TAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILL 157
+KLGL+ H V ILGFNS ENF +AVG + AGG+ GIY T+S EACQY+ N+++
Sbjct: 60 LKGLKQAHSAVAILGFNSPEWFFSAVGTVFAGGIVTGIYTTSSPEACQYIAYDCCANVM 119

Query: 158 VENDOQLOKILSIPOSSLEPLKAIQYRLPM-KNNNLYSWDDFMELGRSIPDTOLEQVI 216
V+ +QL+KIL I L LKA++ Y+ P K N+Y+ ++FMELG +P+ L+ +I
Sbjct: 120 VDTQKQLEKILKI-WKQLPHLKAVVIYKEPPPNKMANVYTMEEFMELGNEVP EALDAII 178

Query: 217 ESQANQCAVLIYTS GTTGIPKGVMLSDNITWIA--GAVTKDFKLT-DKHETVVSYLPL 273

```

++Q+ NQC VL+YTS GTT PKGVMLS DNITW A G+ D + + + E VVSYLPL
Sbjct: 179 DTQQPNQCCVLVYTS GTTGNPKGVMLSQDNITWTARYGSQAGDIRPAEVQQEVVSYLPL 238
Query: 274 SHIAAQMDIWPVPIKIGALTYFAQADALKGTLVSTLKEVKPTVFIGVPQIWEKIHVMVK 333
SHIAAQ+ D+W I+ GA FA+ DALKG+LV+TL+EV+PT +GVP++WEKI E +++
Sbjct: 239 SHIAAQIYDLWTGIQWGAQVCFAPDALKGSLVNTLREVEPTSHMGVPRVWEKIMERIQE 298
Query: 334 NSAKSMGLKKKAFVWNRNIGFKVNSKKMLGKYNTPVSYRMAKTLVFSKVKTSGLDHC 393
+A+S +++K +WA ++ + N G P + R+A LV +KV+ +LG C
Sbjct: 299 VAAQSGFIRKMLLMWMSVTLEQNLG-CFSGDLKPFTRLDYLVLAQRQALGFACQK 357
Query: 394 FISGTAPLNQETAFFFLSDIPIGELYGLSESSGPHITISQNNYRLSLGKILTGCNML 453
G AP+ ET FFL L+I + YGLSE+SGPH +S+ NYRL S GK++ GC+ L
Sbjct: 358 NFYGAAPMMAETQHFFLGLNIRLYAGYGLSETSGPHMSSPYNYRLYSSGKLVPGCRVKL 417
Query: 454 FOQNKDGI GEICLWGRHIFMGYLESETETETEAIDDEGLHSGDLGLDGLGFLYVTGHIK 513
Q+ +GIGEICLWGR IFMGYL E +T EAID+EGWLH+GD G+LD GFLY+TG +K
Sbjct: 418 VNQDAEGIGEICLWGRITIFMGYLNMDKTCEAIDDEGLHSGDAGRLDADGFLYITGRK 477
Query: 514 EILITAGGENVPPIPVETLVKKKIPIISNAMLVGDKLFLSMLLTLCENMQMSGEPLDK 573
E+ITAGGENVPP+P+E VK ++PIISNAML+GD+ KFLSMLLTLC ++ + + D
Sbjct: 478 ELIITAGGENVPPVPIEEAVKMLPIISNAMLIGDQRKFLSMLLTLCPLDPTSDQTDN 537
Query: 574 LNFEAINFCRGLSQASTVTEMVKQDPLVYKAIQGINAVNQEAMNNAQRIEKWVILEK 633
L +A+ FC+ +GS+A+TV+E++++D VY+AI++GI VN A++I+KWLLE+
Sbjct: 538 LTEQAVEFCQVRGSRATTVEIEIEKKDEAVYQATEEGIRRVNMNAAAPYHIQKWAILER 597
Query: 634 DFSIYGGLGPMMLKRRHFVAQKYKKQIDHMY 665
DFSI GGELGP MKLKR V +KYK ID Y
Sbjct: 598 DFSISGGELGPTMKLRLTVLEKYKGIIDSFY 629

```

Pedant information for DKFZphtes3_35kl6, frame 2

Report for DKFZphtes3_35kl6.2

```

[LENGTH] 666
[MW] 74344.97
[PI] 8.67
[ROMOL] TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens
mRNA for KIAA0631 protein, partial cds. 1e-176
[FUNCAT] 1 lipid metabolism [H. influenzae, HI0002] 2e-55
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.13 lipid and fatty-acid transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YER015w]
2e-29
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YMR246w]
2e-23
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YMR246w] 2e-23
[BLOCKS] BL00455
[SCOP] d1lci_ 5.19.1.1.1 Luciferase [Firefly (Photinus pyralis) 1e-49
[EC] 1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolysing) 9e-17
[EC] 6.2.1.3 Long-chain-fatty-acid--CoA ligase 4e-34
[EC] 5.1.1.11 Phenylalanine racemase (ATP-hydrolysing) 6e-08
[EC] 6.2.1.12 4-Coumarate--CoA ligase 8e-18
[PIRKW] duplication 6e-07
[PIRKW] phosphopantetheine 3e-12
[PIRKW] multifunctional enzyme 3e-06
[PIRKW] ligase 6e-08
[PIRKW] acid-thiol ligase 4e-34
[PIRKW] transmembrane protein 5e-22
[PIRKW] monooxygenase 9e-17
[PIRKW] hydrolase 4e-34
[PIRKW] peroxisome 9e-15
[PIRKW] antibiotic biosynthesis 3e-12
[PIRKW] isomerase 6e-08
[PIRKW] flavonoid biosynthesis 1e-17
[PIRKW] magnesium 9e-15
[PIRKW] ATP 5e-22
[PIRKW] oxidoreductase 9e-17
[PIRKW] liver 2e-31
[SUPFAM] alpha-aminoacyl-cysteine-valine synthetase 3e-07
[SUPFAM] human long-chain-fatty-acid--CoA ligase 4e-34
[SUPFAM] gramicidin S synthetase I 6e-08
[SUPFAM] peptide synthetase ppsE 7e-06
[SUPFAM] gramicidin S synthetase I repeat homology 3e-12
[SUPFAM] peptide synthetase ppsD 2e-07

```

```

(SUPFAM)      probable acyl-CoA ligase medium chain 2e-09
(SUPFAM)      acetate--CoA ligase 8e-10
(SUPFAM)      acetate--CoA ligase homology 4e-54
(SUPFAM)      surfactin synthetase 3e-12
(SUPFAM)      4-coumarate--CoA ligase 8e-18
(SUPFAM)      short-chain alcohol dehydrogenase homology 8e-07
(SUPFAM)      acyl carrier protein homology 2e-29
(PROSITE)     MYRISTYL 12
(PROSITE)     AMP_BINDING 1
(PROSITE)     AMIDATION 1
(PROSITE)     CAMP_PHOSPHO_SITE 1
(PROSITE)     CK2_PHOSPHO_SITE 9
(PROSITE)     TYR_PHOSPHO_SITE 3
(PROSITE)     PKC_PHOSPHO_SITE 10
(PROSITE)     ASN_GLYCOSYLATION 2
(PFAM)        AMP-binding enzymes
(KW)          Irregular
(KW)          3D
(KW)          LOW_COMPLEXITY 1.80 %

SEQ  MTGTPKTOEGAKDLEVDNMNKEVTPRLWTTTCRDGEVLLRLSKHGPGHETPMTIPEFFRES
SEG  .....
11ci- .....

SEQ  VNRFQTYPALASKNGKKWEILNFQYYEACRKAASLIKLGLERFHGVGILGFNSAEWFI
SEG  .....
11ci- .....

SEQ  TAVGATLAGGLCVGIATNSAEACQYVITHAKVNILLVENDQQLKILSIPOSSLEPLKA
SEG  .....
11ci- .....

SEQ  IIQYRLPMKKNNNLYSWDDFMELGRSIPDTQLEQVIESQANQCAVLIYTSGETTGIPKGV
SEG  .....
11ci- .....

SEQ  MLSDNITWIAGAVTKDFKLTDKHETVVSYLPLSHIAQMMDIWVPIKIGALTYFAQADA
SEG  .....
11ci- .....

SEQ  LKGTIVSTLKEVKPTVFIGVPQIWEKIHEMVKNKSAKSMGLKKKAFVWARNIGFKVNSKK
SEG  .....
11ci- .....

SEQ  MLGKYNTPVSYRMARTLVFSKVKTSGLDHCFSISGTAPLNQETAEFFLSLDIPIGELY
SEG  .....
11ci- ..... TTTTCEETTTTCCCHHHHHHHHHCCCCBCEE

SEQ  GLSESSGPHTISNQNNYRLSCGKILTGCKNMLFQONKDGIGEICLMGRHIFMGYLESET
SEG  .....
11ci- ..... ECGGGTTEEECCCCCEEEETTTTCEETTEEEETTTTCEETTTTCEETTTTHH

SEQ  ETTEAIDDEGWLHSGDLGQLOGLGFLYVTGHINEILITAGGENVPPIPVETLVKKKIPII
SEG  .....
11ci- ..... HHHHHBTTTCEEEEEETTTTCEEE-----ECEETTEECHHHHHHHHHHT-TTE

SEQ  SNAMLVGDKLKLFLSMLLTLCENMNSGEPDLKLNFEAINFCRLGSOASTVTENVKQOD
SEG  .....
11ci- ..... EEEEEEE.....

SEQ  PLVYKAIQQGINAVNQEAMNNAQRIEKWVILEKDFS IYGGELGPMMLKRHFVAQYKKQ
SEG  .....
11ci- .....

SEQ  IDHMYH
SEG  .....
11ci- .....

```

Prosites for DKFZphtes3_35k16.2

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	246->250	ASN_GLYCOSYLATION	PDOC00001
PS00004	332->336	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005

PS000005	308-->311	PKC_PHOSPHO_SITE	PD0C00005
PS000005	335-->338	PKC_PHOSPHO_SITE	PD0C00005
PS000005	358-->361	PKC_PHOSPHO_SITE	PD0C00005
PS000005	370-->373	PKC_PHOSPHO_SITE	PD0C00005
PS000005	558-->561	PKC_PHOSPHO_SITE	PD0C00005
PS000006	30-?34	CK2_PHOSPHO_SITE	PD0C00006
PS000006	52-->56	CK2_PHOSPHO_SITE	PD0C00006
PS000006	173-->177	CK2_PHOSPHO_SITE	PD0C00006
PS000006	196-->200	CK2_PHOSPHO_SITE	PD0C00006
PS000006	206-->210	CK2_PHOSPHO_SITE	PD0C00006
PS000006	210-->214	CK2_PHOSPHO_SITE	PD0C00006
PS000006	308-->312	CK2_PHOSPHO_SITE	PD0C00006
PS000006	478-->482	CK2_PHOSPHO_SITE	PD0C00006
PS000006	591-->595	CK2_PHOSPHO_SITE	PD0C00006
PS000007	659-->666	TYR_PHOSPHO_SITE	PD0C00007
PS000007	658-->666	TYR_PHOSPHO_SITE	PD0C00007
PS000007	597-->605	TYR_PHOSPHO_SITE	PD0C00007
PS000008	3-->9	MYRISTYL	PD0C00008
PS000008	65-->71	MYRISTYL	PD0C00008
PS000008	124-->130	MYRISTYL	PD0C00008
PS000008	130-->136	MYRISTYL	PD0C00008
PS000008	134-->140	MYRISTYL	PD0C00008
PS000008	235-->241	MYRISTYL	PD0C00008
PS000008	235-->245	MYRISTYL	PD0C00008
PS000008	303-->309	MYRISTYL	PD0C00008
PS000008	387-->393	MYRISTYL	PD0C00008
PS000008	421-->427	MYRISTYL	PD0C00008
PS000008	498-->504	MYRISTYL	PD0C00008
PS000008	586-->592	MYRISTYL	PD0C00008
PS000009	74-->78	AMIDATION	PD0C00009
PS00455	227-->239	AMP_BINDING	PD0C00427

HMM_NAME		AMP-binding enzymes	
HMM		*TYRELNERANRLARLHRS+GIGrPGDlVgIMMDRSMWMIvAMLGWKAG ++ E + + L + + G VGI+ S + + + G + AG	
Query	82	NFNQYYEACRAAKASLI-KLGLERFHGVGILGFNSAEWFITAVGAILAG	129
HMM		GAYVIPIDeYPDeIRqYMLEDSGArLLITOrh... HmqRIpmwvvvdH + E QY++ ++ + L+++ + + I P+++++	
Query	130	GLCVGIYATNSAEACQYVITHAKVNILLIVENQQQLKLSIPQSSELEPK	179
HMM		Iiviiwve..... WddlWHHeeeNppqwPdeDLAYIIV +I++ + + + + + E + + + + A + IY	
Query	180	AIIQYRLPMKKNNNLYSWDFFMELGRSIPDTOLEEVIESQRKANVLIV	229
HMM		TSGTGTGPKGMVIEHrINIVnqCmWNRYMeedDRILMFITSdpYWfDA TSCTTG PKGMV++H NI+ + + + + T+ + + + + + + A	
Query	230	TSGTGTGIPKSLSHDNITIAGAVTKDFKLTDKHETVSYSLE-LSHIAA	278
HMM		SVMDMFWPLLnGaTLyIpPeEtRdPerWmQYIRHgITWylTPSMFRM +++D+ P+ GA Y+ + + + + + + + T+ + + P + + +	
Query	279	QMMDIWPIKIGALTYFAQADAL--RGTLVSTLKVKPTVFVGVPQIEWK	326
HMM		Lmpdi..... + +	
Query	327	IHEMVKNsAKSMLKKKAFVWARNIGFKVNSKMMLGKYNTFPVSrMAKT	376
HMM	psLRhvmFGepLSPehmdWWrkfrfgkgrINIHWYT + + + + + G PL+E++ + + + + + IY + +	
Query	377	LVFSKVKTSGLGDHCFSFISGTAPLQQTAEFFL-SLD-IPIGELYGLS	423
HMM		ETTvtTWmrIIdpepeq+rwiPIGRPiPntqWYImDDnmMQIQiGviGE ++ + + + T + + + + + G + N G IGE	
Query	424	ESSGFHTISNQNN--Y---RLSSCGRILOTCKNNMLFQN---KDG-IGE	463
HMM		LYLGWGPGVARGYNNRPPELTzeipNPFWPGEYrGrWNrRZGTOLAR ++ G + + + GY+ + T E+ + + + + + GDL+ +	
Query	464	CICGW-RHIFMGYLESETETTEIDADEGG-----LHSGDLOQ	499
HMM		wLPdgNIeYLGRID.DQVKYIKRIELGEIehqLr.qHPgiqeEAVV+ + G++ G I + + G++ + + + E+ + + F+ I+ A	
Query	500	LDGLGFLPYTHGIKEILLTAGGENVPPIVETLKKKKIIPISNAML	545

DKF2phtes3_35k24

group: transmembrane protein

DKF2phtes3_35k24 encodes a novel 514 amino acid protein without similarity to known proteins.

The novel protein contains 5 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown ;

membrane regions: 5

Summary DKF2phtes3_35k24 encodes a novel 514 amino acid protein. No homologues found in bacteria yeast and C.elegans, specific for mammals?

unknown

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2706 bp

Poly A stretch at pos. 2696, polyadenylation signal at pos. 2675

```
1 CCGTGTGCAG TCGCCCGCG CCGCCGCGA CCCTTCGGGT AAACACGAA
51 CTGGGAGTTC TGAAGAATGG GTAAAGACTT TCGTACTAT TTCCAGCATC
101 CCTGGTCTCG CATGATTGTG GCTTACTTGG TGATCTTCTT TAACCTCTTA
151 ATATTTCGGG AGGACCCAGT TTCTCATAGC CAAACAGAAG CCAATGTTAT
201 TGTGTGTGGA AACTGTTTTT CATTGTGTAC AAATAAATAC CCTAGAGGAG
251 TTGGCTGGAG GATTTGAAG GTGCTTCTAT GGCTACTTGC CATTCTCACA
301 GGACTAATAG CTGGCAAAAT TCTGTTCAT CAGCGTTTGT TTGGTCAGTT
351 GCTCCGATTA AAAATGTTTC GAGACATCA TGGGTCTGGG ATGACAAATG
401 TCTTCAGCAC AATCTCTTTT CTCTTCATAT TTCTTCACAT ATACAAACAG
451 ATTCTTCTAA TGGATGGGAA CATGGGAGCA TATATCATT CAGACTATAT
501 GGGCATCCGA AATGAAGTT TCATGAATTT AGCTGCAGTA GGGACCTGGA
551 TGGGGGACTT TGTCAAGCTT TGGATGGTCA CTGATATGAT GCTTCAGGAC
601 AAACCCATAT CTGACTGGGG AAAATCAGCA AGAGCTTTCT GGAAGAAAGG
651 AAATGTTAGG ATCACTTTAT TCTGGACAGT TCTTTTACT CTGACGTCTG
701 TGGTTGTACT TGTGATTACA ACGGACTGGA TCAGCTGGGA CAAGCTGAAT
751 CGGGGATTTT TGCCCACTGA TGAAGTTTCC AGAGCATTCC TTGCTTCTTT
801 TATCTTGGTC TTTGACCTTC TTAATTGTAT GCAGGACTGG GAATTCCTAC
851 ATTTTCATGG AGATGTTGAT GTAAATCTCC CTGGTTTGCA CACCCCTCAC
901 ATGCAGTTCA AGATCTCTTT CTCCAGAAAT ATCTTCAAGG AGGAATATCG
951 TATTCAATAA ACAGGCAAT GGTTAATAA TGGAAATATC TTCTCTGTCT
1001 TGATTTTGGG TCTTAATATG TGGAGAACC AAATATTTTA TAAACCTCAT
1051 GAATATGGGC AATATATCGG CCGGGGCGAG AAGATATATA CAGTGAAAGA
1101 CTCAGAAAGT TTAAGAATTT TGAACAGAAC CAAGCTATCC TGGGAATGGA
1151 GGTCCAATCA CACTAACCTT CGGACTAATA AAACATATGT TGAGGGAGAC
1201 ATGTTCTTAC ACAGCAGGTT CATAGGAGCC AGTCTTGATG TCAAGTGTCT
1251 GGCCTTTGTT CCAAGCCTGA TAGCCTTTGT GTGGTTTGGG TTCTTTATTT
1301 GGTTCCTTGG ACGATTTTGG AAAAATGAGC CACGCATGGA GAATCAAGAC
1351 AAAACTTACA CTCCGATGAA AAGAAAATCT CCATCAGAAC ATAGCAAGAA
1401 CATGGGAATC ACTCGAGAAA ACACCCAGGC TTCAGTAGAA GACCCCTTGA
1451 ATGACCCCTT TTTGGTTTGC ATCAGGCTGT ACTTCAATGA GATCGTCTAC
1501 AAGTCTTCCC ACCTAACCTC GGAAAGCTTG AGCTCACAGT TGAAGCAATC
1551 TACTAGTGCA ACAGAGCTGT ATCAAGACCC AACGACTTCT AAAAGTACAC
1601 CTACGAACTA GACTCGGAGA TAGACTTGGG GATAACACAA AAAGCAACCT
1651 TGAGCTAAC TTTAAAAATT TAGCTTTTCC TTTTGTATAT GTAAGGTTTA
1701 CGTAGTGTAA GGTAAAAAAT TGAACAATGC CACAACGGTG CTCACATGCG
1751 TTTTCTTAGG ATTCAATTGT TTCTATTGTT ATTATAATAC ACGTGCCTAC
1801 TGTAATACTA ACAGTCTCTT AGAGATTGCT TTTCACAATT GCACAGCTAC
1851 TTACTGACTT TACAGCATAG TGGAGATTA GCTGATGACC CATGTATCTG
1901 ATGTTCAACC ATAGTGGTGC CTGAGACAT TAAACTGTTT TTAAGTGTAC
1951 CAGAAATGAA GTGTGGAACA GTTACCTAAC CTATTTCACA TGGGCGTTTT
2001 GTATACAAC TTTTGTATCT ACACCTGATG TCTGAGCAGA AAACAGAAAT
2051 AGCTAAATGT GACTCAGGAA GTATCTCTTG GTTTCTTATT CAGCAGCAGA
2101 GTTGGTGACT TTGACAACTG GACTGCAGAG AAACATGGTG ATCACCCTTT
2151 AATTTTTTAT GGCCTGTCTG CAATATATAA TACAGATGCA AATATCAGTA
2201 ATAGGAGATC CATACCCCAA CATGGGTCAC TACTCTGTGA ATGTGACTTT
2251 CTCCACACAG TAATTGAAT GAGGTGATGA TACCTAATTA TGTTTTCTTA
2301 ATTAAGATA AATTGCTACT TGATTAAAA TCCTGCCCTT CACCTTTGGG
```


PCT/IB00/01496

2351	AACAAAGGTT	AAGAGACACA	GTGGGGCGAA	CTCTCAAAAT	TATTGGCAAT
2401	CAGCAAAAGT	CCGACGACAA	CAGACAGCAT	AGGTTTTCAT	CATTAGAGAG
2451	CAGATCTGAT	TCAATGCTGT	TATTCGCTGT	TATTTCTGCT	CAAAATATGC
2501	CTCTGGATAT	GAAAATTTGA	AAACATATGC	CAACCTGTGC	CAGGGAAGAT
2551	CTCTCAAAAA	TTATGACGCG	GAACTTCTGC	AGGTGAGAGA	CGCGGTGATG
2601	AAAGAATTGT	CTAATCGTGT	TGTTTTCGCT	ATGTGTTTGT	TGTTTGTGTT
2651	TTTTAAGAAC	TAAATATTGC	ACATTAAATA	ATAAGAATTA	TACAGCAAAA
2701	AAAAAA				

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 67 bp to 1608 bp; peptide length: 514
Category: putative protein

1	MGKDFRYKYP	PPMSRMIVAY	LVIFENFLTG	ADGVPVSHQT	EANVIVVYGR
51	PDGKYYNKKP	LVGKRLKVL	LWLAIFLTLF	IACKFLFHQR	LFQGLLRLKM
101	PRDGHGSHMT	MFSFSLITFL	IFSHIYNTPL	LDGMGGNMYA	ITDYMKNRNE
151	SFKMLAASVT	MWGDVEFTAW	VDHMLQDQRT	YPWQKSGSARA	YFAGKGNVIT
201	LFMTLVLTFL	SVSVLLVITD	WISWDLKRLN	FLPDSVBSRA	FLASFLIVTD
251	LVTLVMDQEF	PHMGDQDVON	WISHLTPHMQ	KPIFFKQFPI	KEEYIRHITG
301	KWNVGSLIFL	VLTLDLNMHK	NOIFPKRHEY	GOITFGPQKI	YWKDSGSL
351	DLNRGKLSWE	WMLVGLVQ	QWVWVWVW	YVWVWVWV	YVWVWVWV
401	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV
451	ADQTSVDEP	LDNPSLVCIR	SDNEIYVYKS	SLHTSENLSL	QLWNETSATE
501	ENPORTSSKS	TPFN			

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_35k24, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35k24, frame 1

Report for DKFZphtes3_35k24.1

[LENGTH]	514	
[MW]	60185.03	
[pI]	8.67	
[PROSITE]	MYRISTYL	5
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	8
[PROSITE]	TYR_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	7
[PROSITE]	ASN_GLYCOSYLATION	6
[KW]	SIGNAL PEPTIDE	32
[KW]	TRANSMEMBRANE	5
[KW]	LOW COMPLEXITY	15.37

[illegible]

[illegible]

Prosite for DKFZphtes3_35k24.1

PS000001	149->153	ASN_GLYCOSYLATION	PDOC000001
PS000001	353->357	ASN_GLYCOSYLATION	PDOC000001
PS000001	364->368	ASN_GLYCOSYLATION	PDOC000001
PS000001	371->375	ASN_GLYCOSYLATION	PDOC000001
PS000001	487->491	ASN_GLYCOSYLATION	PDOC000001
PS000001	493->497	ASN_GLYCOSYLATION	PDOC000001
PS000004	435->439	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	55->58	PKC_PHOSPHO_SITE	PDOC000005
PS000005	187->190	PKC_PHOSPHO_SITE	PDOC000005
PS000005	299->302	PKC_PHOSPHO_SITE	PDOC000005
PS000005	342->345	PKC_PHOSPHO_SITE	PDOC000005
PS000005	347->351	PKC_PHOSPHO_SITE	PDOC000005
PS000005	370->373	PKC_PHOSPHO_SITE	PDOC000005
PS000005	507->510	PKC_PHOSPHO_SITE	PDOC000005
PS000006	38->42	CK2_PHOSPHO_SITE	PDOC000006
PS000006	342->346	CK2_PHOSPHO_SITE	PDOC000006
PS000006	348->352	CK2_PHOSPHO_SITE	PDOC000006
PS000006	373->377	CK2_PHOSPHO_SITE	PDOC000006
PS000006	438->442	CK2_PHOSPHO_SITE	PDOC000006
PS000006	456->460	CK2_PHOSPHO_SITE	PDOC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC000006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC000006
PS000007	326->344	TYR_TYR_PHO_SITE	PDOC000007
PS000008	48->55	MYRISTYL	PDOC000008
PS000008	79->85	MYRISTYL	PDOC000008
PS000008	106->112	MYRISTYL	PDOC000008
PS000008	134->140	MYRISTYL	PDOC000008
PS000008	159->165	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphtes3_35k24.1)

DKFZphtes3_35n12

group: metabolism

DKFZphtes3_35n12 encodes a novel 315 amino acid protein with strong similarity to ADP,ATP carrier T (ANT) proteins.

The novel protein contains three mitochondrial energy transfer signatures and is closely related to the ADP/ATP translocator, or adenine nucleotide translocator (ANT), a protein most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.

The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

strong similarity to ADP/ATP carrier proteins

EST hits to mouse and drosophila

Sequenced by DKFZ

Locus: unknown

Insert length: 1803 bp

Poly A stretch at pos. 1793, polyadenylation signal at pos. 1772

```
1 AGCGTCCCAA GAGCCACTTT CTCGCCAGTA CGATGCTGCA CCGGTTTTCC
51 GGTITTCGCG TTCCCTTCAT CGTAGCTCCC GTACTCATTT TTAGCCACATG
101 CTGCGCGGTT TTATATCCCT CTCATCATG CATCGTGAGC CTGCGAAAAA
151 GAAGCGAGAA AAGCGGCTGT TTGACGCCCTC ATCCTTCGGG AAGGACCTTC
201 TGGCGCGCGG AGTCGCGGCA GCTGTGTCCA AGACAGCGGT GCGGCCCATC
251 GAGCGGGTGA AGCTGCTGCT GCAGGTGCGA GCGTCGTCGA AGCAGATCAG
301 CCCCAGGGCG CGGTACAAAG GCATGGTGGA CTGCTGGTG CGGATTCCCTC
351 GCGAGCAGGG TTTCTTCAGT TTTTGGCGTG GCAATTTGGC AAATGTTATT
401 CGGTATTTTC CAACACAAGC TCTAAACTTT GCTTTTAAGG ACAAATACAA
451 GCAGCTATTC ATGCTCGGAG TTAATAAAGA AAAACAGTTC TGGAGGTGGT
501 TTTTGGCAAA CCTGGCTTCT GGTGGAGCTG CTGGGGCAAC ATCCTTATGT
551 GTAGTATATC CTCTAGATTT TGCCCGAACC CGATTAGGTG TCGATATTGG
601 AAAAGGTCTT GAGGAGCGAC AATTCAGGG TTTAGTGAC TGATATTGSA
651 AATAGCAAAA ATCAGATGCA ATTGCTGGTT TATACCAAGG GTTTGGTGTT
701 TCAGTACAGG GCATCATTTG GTACCGAGCC TCTTATTTG GAGCTTATGA
751 CACAGTTAAG GGTTTATTAC CAAAGCCAAA GAAAACTCCA TTTCTTGCT
801 CCTTTTCAT TGCTCAAGTT GTGACTACAT GCTCTGGAAT ACTTCTTAT
851 CCCTTTGACA CAGTTAGAAG ACGTATGATG ATGCAGAGTG GTGAGGCTAA
901 ACGGCAATAT AAAGGAACCT TAGACTGCTT TGTGAAGATA TACCAACATG
951 AAGGAATCAG TTCTTTT TTTTGGCGCT TCTCCAATGT TCTTCGCGGT
1001 ACAGGGGGTG CTTTGGTGT GGTATTATAT GATRAAATTA AAGAATCTT
1051 TCATATTGAT ATTGGTGSTA GGTAAATCGG AGAGTAAATT AAGAAATAAC
1101 ATGGATTFAA CTTGTTAAAC ATACAAATTA CATAGCTGCC ATTTCATAC
1151 ATTTTGATAG TGTTATTGTC TGATTTTGT TAAAGTGCTA GTTCTGCAAT
1201 AAAGCATACA TTTTTCAGC AATTTAAATA CTAAAAATCA GATAAATGTG
1251 GATTTTCCTC CCACTTAGAC TCAAAACAT TTTAGTGTGA TATTTCATTT
1301 ATTATAGGTA GTATATTTTA ATTTGTAGT TTAATAATCT TTTTATGATT
1351 AAAAATTAAT CATATAATCC TAGATTATG CTGAAATCTA GGAAATGAAA
1401 GTAGCGTCTT TTAATTGCT ATTCAATTAA TATACCTGTT TTCCCATCTT
1451 TTGAAGTCAT ATGGTATGAC ATATTCTTAA AAAGCTTATC AATAGATGTC
1501 ATCATATGTG TAGGCAGAAA TAAGCTTTGT TCTATATCTC TTCTAAGACA
1551 GTGTATTATA CTGGTATATA TATTTCAGT ATCAGCCTTT GATTATAGAT
1601 GTGATCATTT AAAATTGAT AATGACTTTA GTGACATTAT AAAACTGAAA
1651 CTGGAAAAATA AAATGGCTTA TCTGCTGATG TTTATCTTTA AAAATAATAA
1701 AATCTTGCTA GTGTGAATAT ATCTTAGAAC AAAAGGTATC CTCTTGAAAA
1751 TTAGTTTGTA TATTTTGTG ACAATAAAGG AAGCTTAAC TTTAAAAAAA
1801 AAA
```

BLAST Results

No BLAST result

Medline entries

96289608:
Molecular biological and quantitative abnormalities of
ADP/ATP carrier protein in cardiomyopathic hamsters.

Peptide information for frame 2

ORF from 128 bp to 1072 bp; peptide length: 315
Category: strong similarity to known protein
Classification: Metabolism
Prosite motifs: MITOCH_CARRIER (40-50)
MITOCH_CARRIER (145-155)
MITOCH_CARRIER (242-252)

1 MHREPARKKA EKRLFDASSF GKDLLAGGVA AAVSKTAVAP IERVKLLQV
51 QASSKQISPE ARYKGMVDCL VRIPREQOFF SFWRGNLANV IRYFPTQALN
101 FAFKDYKQL FMSGVNKEKQ FWRFLANLA SGGAGATSL CVVPLDFAR
151 TRLCVDIGKG PEERQFKGLG DCIMKIAKSD GIAGLYQGFV VSVQGIIVYR
201 ASYFGAYDTV KGLLPKPKKT PFLVSFFIAQ VVTTCSGILS YPFDTVRRRM
251 MQSGGEAKRQ YKGTLDCEVK IQHEGSISS FRGAFSNVLR GTGGALVLLV
301 YDKIKEFFHI DIGGR

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n12, frame 2

PIR:S37210 ADP,ATP carrier protein T1 - mouse, N = 1, Score = 1127, P = 2.7e-114

PIR:A44778 ADP,ATP carrier protein T1 - human, N = 1, Score = 1125, P = 4.4e-114

TREMBL:DNADPATPT_2 product: "ADP/ATP translocase"; Drosophila melanogaster gene encoding ADP/ATP translocase, N = 1, Score = 1124, P = 5.6e-114

PIR:XWBO ADP,ATP carrier protein T1 - bovine, N = 1, Score = 1121, P = 1.2e-113

>PIR:S37210 ADP,ATP carrier protein T1 - mouse
Length = 298

HSPs:

Score = 1127 (169.1 bits), Expect = 2.7e-114, P = 2.7e-114
Identities = 214/293 (73%), Positives = 248/293 (84%)

Query: 17 ASSEKGLLAGGVA AVSKTAVAPIERVKLLQVQASSKQISPEARYKGMVDCLVRIPRE 76
A SF KD LAGG+AAAVSKTAVAPIERVKLLQVQ+SKQIS E +YKG+DC+VRIP+E
Sbjct: 5 ALSFLKDLFAGGIAAAVSKTAVAPIERVKLLQVQHASKQISAEKQYKGIIDCVVRIPKE 64
Query: 77 QGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQF+GV++KQFWR+F NLASGGAAG 136
QGF SFWRGNLANVIRYFPTQALNFAFKDKYKQF+GV++KQFWR+F NLASGGAAG
Sbjct: 65 QGFLSFWRGNLANVIRYFPTQALNFAFKDKYKQIFLGGVDRHKQFWRVAGNLASGGAAG 124
Query: 137 ATSLCVVYPLDFARTRLCVDIGKGPPEERQFKGLGDCIMKIAKSDGIAGLYQGFVSVQGI 196
ATSLC VYPLDFARTRL D+CKG +R+F GLGDC+KI KSDG+GLYQGF VSVQGI
Sbjct: 125 ATSLCFVYPLDFARTRLAADVCKGSSQREFNGLGDCILTKIFKSDGLKGLYQGFVSVQGI 184
Query: 197 IVYRASVYFGAYDTVKGLLPKPKKT PFLVSFFIAQVVTTCSGILSYFPDTRRRMMQSGE 256
I+YRA+YFG YDT KG+LP PK +VS+IAQ VT +G++SYFPDTRRRMMQSGE
Sbjct: 185 IIVRAAYFGVYDTAKGMLPDPKNVHIIVSWMIAQSVTAVAGLVSYFPDTRRRMMQSGR 244
Query: 257 --AKRQYKGTLDCEVKIQHEGSISSFRGAFSNVLRGTGGALVLLVYDKIKEF 307
A Y GTLDC+KI +EG ++FF+GA+SNVLRG GGA VLVLVD+IK++
Sbjct: 245 KGADIMYGTLDCEWRKIAKUEGANAFFGAWSNVLRGAGAFVLLVYDEIKKY 297

Pedant information for DKFZphtes3_35n12, frame 2

Report for DKFZphtes3_35n12.2

[LENGTH] 315

[illegible]

Prosites for DKFZphtes3_35n12.2

PS00215	40->50	MITOCH_CARRIER	PDOC00189
PS00215	145->155	MITOCH_CARRIER	PDOC00189
PS00215	242->252	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphtes3_35n12.2

HMM_NAME	Mitochondrial carrier proteins	
HMM	*pFwkdFLAGGIAGMeHTvMFPIDtIKTRMOLQgEMpH...shpRYkGMI +F+KD+LAGG+A++++T+++PI+++K+++Q+O +++ RYKGM+	
Query	19 SFGKDLLAGGVAAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMV	67
HMM	dCFRwIwkNEGWRLWRGLgANvIRYIPgWaIRFGFYEFMKEMFiDyIge DC+ +I++++G++++WRG++ANVIRY+P+++F+F++ +K +F + +++	
Query	68 DCLVRI PREQGFSSFWGNLANVIRYFPTQALNFAFKDKYKQLFMVGYNK	117
HMM	ddnyWmWfMnYMaGsmAGEwisIitYPMWvVKTRLQaDqkHphsQp.R ++W+WF+ N+++G++AG ++S+ ++YP++++TRL D +++++ R	
Query	118 EKQFWRWFLANLASGGAAG-ATSLCVVYPLDFARTRLGVD--IGKGPZER	164
HMM	hYNGvWncWkIYReEGgFkGLYRGWtPTWNRMI PYqmiYFivYEtLKeW +++G+ +C KI +++G ++GLY+G++ +++++I+Y++ YF++Y+T K +	
Query	165 QFKGLGDCIMKIAKSDG-IAGLYQGFGVSQGIIVYRASYPGAYDTVKGL	213
HMM	lynYtgYnPgprelCMddsPwWhwLIgWmIAGMiaWivSYPFdVVRTRMM L +++ + ++++++I++ ++ +++++I+SYPFd+VR+RMM	
Query	214 LP-----KPK--KTPFLVSFFIAQVVT-TCSGILSYPFdTVRRMM	251
HMM	Mdam.edhkYqSmLDcWmqIYKneGfKGFwKGFwPRIMRiMPWtAIMPmi M+S+ +++++Y+++LDC+++IY++EG+ +F++G+ +RR+ ++A++++	
Query	252 MQSGEAKROYKGTLDCEVKIYQHEGISFFRGAfSNVLRGT-GGALVLVL	300
HMM	YEqMKwFL* Y+ +K+F+	
Query	301 YDKIKEFF	308

DKFZphtes3_35n24

group: testes derived

DKFZphtes3_35n24 encodes a novel 365 amino acid protein without similarity to known proteins.

The novel protein contains a Prosite Ig(Immunoglobulin)-MHC pattern. This pattern represents a domain, approximately one hundred amino acids long and including a conserved intra-domain disulfide bond (Ilg domain1). Thus, the novel protein is a new member of the Ig-superfamily. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1589 bp

Poly A stretch at pos. 1579, polyadenylation signal at pos. 1560

```
1 CGATCGTCAC GTGACGCCGG GGTTCAGCGT ATCCTTGCTG GGCAACCGTC
51 TTAGAGACCA GCACTGCTGG CTGCACCATG AATGTGATCT ACCCACTGGC
101 AGTCCCAAG GGGCGCAGAC TCTGCTGTGA GGTGTGGGAA GCGCCAGCCG
151 AGCGGGTGTG CGCGGCCTGC ACAGTCACCT ATTACTGTGG GGTGGTACAT
201 CAGAGGGCTG ACTGGGACAG CATCCATGAG AAAATATGTC AGCTCTTGAT
251 TCCACTGGCG ACTTCCATGC CCTTCTACAA TTCAGAGGAA GAACGGCAGC
301 ATGGCCCTGA CGAGCTGCAG CAGCGGCAGA AGTATTTGAT TGAATCTGTC
351 TACACCATAG CCCAGAAATA CCTCTTTGAA GGGAAACACG AAGATGCTGT
401 ACCAGCAGCT TTGCAGTCCC TTGCTTCCG TGTGAAGCTG TATGGCCTGA
451 GCTCCGTAGA GCTTGTGCTT GCTTACCCGC TGTGGCCGGA GCGCCAGCCTT
501 GGTCTGGGCC GAATCGTTCA GGCTGAAGAA TATCTATTCC AAGCCAGCTG
551 GACAGTCCTC AAATCAACTG ACTGTAGTAA TGCCACCCAC TCTTTACTGC
601 ATCGGAATCT GGGACTTCTC TATATAGCTA AGAAAACTA TGAAGAGGCC
651 CGTTATCATC TGCCCAATGA TATTTATTTT GCCAGTTGTG CATTGGGAAC
701 AGAGGACATT AGGACTTCAG GAGGCTACTT CCACCTGGCT AATATATTCT
751 ATGACCTTAA AAAGTTGGAC CTGGCAGACA CATTGTACAC CAAGGTCTCT
801 GAGATCTGGC ATGCATATTT GAACAATCAC TATCAAGTCC TCTCACAGGC
851 TCACATCCAA CAAATGGATT TACTGGGCAA ACTATTTGAC AATGACACTG
901 GCTTGGATGA AGCCCAAGAA GCAGAAGCCA TTCGCATCCT GACTTCAATC
951 TTGAACATTC GAGAATCTAC ATCTGACAAA GCGCCCAAAA AAACCATCTT
1001 TGTTCTGAAG ATCCTGGTCA TGCTTTACTA CCTGATGATG AATTCTTCAA
1051 AGGCACAGGA ATATGGCATG AGGGCCCTCA GTCTAGCCAA AGACAAACG
1101 CTTGATGTCC ATGAGCAAGC CACCATTCAA GAGTTATTA A GTCTATTTC
1151 AACTGAGAC CATCCCATTA CTTAGTGACC CATGAGCTCT GCATCAAGGG
1201 TTATTCCAGG GGCTACTGAA GATCTAATAT ATTCCAGCCT TGCACAACGT
1251 CTTTGAGGTA CTGTAGACTG CTGAAGTTTC CACCCTTTC CCCTGGGATT
1301 GCACACATAG CTGTTATTTT TTTCTTACAC AGCATATTAA GCGAATATAA
1351 AGCTTTAGGC ATAGAAATCA CTAAAAACGT TGTTTGTCAT GACCTTTGTA
1401 CTTGATTTAT CATGACTTTG TATGACTGAG TAATATGTAG TCAGATCACT
1451 AATATGGTAT TTGTAATTAA ACTACAAATA GTTTGTGATT TCCCAGAAGT
1501 CTTCCAACGA TGCTATGTTT ATACACTTTT GCTAAAGGAG GGGTAAAGGA
1551 GGGGGTAGGG AATAAAGCTA TATTGGAACA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 78 bp to 1172 bp; peptide length: 365
Category: putative protein

Prosite motifs: IG_MHC (35-42)

```
1 MNVIYPLAVP KGRRLCCEVC EPAERVCAA CTVYYCGVV HOKADWDSIH
51 EKICQLLIPL RTSMPFYNSE EERQHLQOL QORQKYLIEF CYTIAQKYLIF
101 EGKHEDAVPA ALQSLAFRVK LYGLSSVELV PAYPLAEAS LGLGRIVQAE
151 EYLFQAQWTV LKSTDCSNAT HSLLRNLGL LYIAKKNYEE ARYHLANDIY
201 FASCAFGTED IRTSGGYFHL ANIFYDLKGL DLADTLTKV SEIWHAYLNN
251 HYQVLSQAH IQQMDLLGKLF ENDTGLDEAQ EAEAIRILTS ILNIRESTSD
301 KAPQKTI FVL KILVMLYYLM MNSSKAQYEG MRALSLAKEQ QLDVHEQSTI
351 QELLSLISTE DHPIT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n24, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35n24, frame 3

Report for DKFZphtes3_35n24.3

```
[LENGTH] 365
[MW] 41768.24
[pI] 5.82
[BLOCKS] BL06273 Heat-stable enterotoxins proteins
[PROSITE] MYRISTYL 1
[PROSITE] IG_MHC 1
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] TYR_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 3
[PROSITE] ASN_GLYCOSYLATION 3
(KW) Alpha_Beta
(KW) LOW_COMPLEXITY 4.11 %

SEQ MNVIYPLAVPGRRLCCEVCEAPAERVCAACTVYYCGVVHOKADWDSIHEKICQLLIPL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RTSMPFYNSEERQHLQOLQORQKYLIEFCYTIQKYLIFEGKHEDAVPAALQSLRFRVK
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LYGLSSVELVPAYPLAEASLGLGRIVQAEYLFQAQWTVLSTDCSNATHSLLRNLGL
SEG .....
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LYIAKKNYEEARYHLANDIYFASCAFGTEDIRTSGGYFHLANIFYDLKGLDLADTLTKV
SEG .....
PRD eeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ SEIWHAYLNNHYQVLSQAH IQQMDLLGKLFENDTGLDEAQEAEAIRILTSILNIRESTSD
SEG .....
PRD hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ KAPQKTI FVLKILVMLYYLMNNSKAQYEGMRALSLAKEQQLDVHEQSTIQELLSLISTE
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DHPIT
SEG .....
PRD ccccc
```

Prosite for DKFZphtes3_35n24.3

PS00001	168->172	ASN_GLYCOSYLATION	PD0C00001
PS00001	272->276	ASN_GLYCOSYLATION	PD0C00001
PS00001	322->326	ASN_GLYCOSYLATION	PD0C00001
PS00005	114->117	PKC_PHOSPHO_SITE	PD0C00005
PS00005	299->302	PKC_PHOSPHO_SITE	PD0C00005
PS00005	323->326	PKC_PHOSPHO_SITE	PD0C00005

WO 01/12659

PCT/IB00/01496

PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	69->73	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00007	85->93	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	185->194	TYR_PHOSPHO_SITE	PDOC00007
PS00008	275->281	MYRISTYL	PDOC00008
PS00009	11->15	AMIDATION	PDOC00009
PS00290	35->42	IG_MHC	PDOC00262

(No Pfam data available for DKF2phtes3_35n24.3)

DKFZphtes3_35n9

group: metabolism

DKFZphtes3_35n9 encodes a novel 607 amino acid protein which is a splice variant of human carboxylesterase (EC 3.1.1.1).

The novel protein contains both, one carboxylesterase B1 and one B2 pattern. In comparison to EC 3.1.1.1, DKFZphtes3_35n9 shows a N-terminal extension and aa 458-474 are missing.

The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

carboxylesterase, splice variant

5' extension of mRNA and N-terminal elongation of protein (64 aa), missing exon! aa 458-474 of JC5408 are missing

Sequenced by DKFZ

Locus: unknown

Insert length: 2888 bp

Poly A stretch at pos. 2878, no polyadenylation signal found

```
1 CTCGGCCTGA GGTGCGAGAG AAGCGGTGAC CGCGGCCCTG GCTGCTCGGA
51 CCGGGGAACA TGATGGTTCG TGGAGCAGAA GGCCTGAGA AGGGACCAAG
101 GCGGCGCTGG GTCGTGCGAG CCAGTAGCGG GCTGAAACGT AGAGGCCAGA
151 ACCAGGTCTC AGGGGGCACT AAAGGGCGTC GGAGGTAAATC CCCACACCGC
201 TTCTCTCTTG AGTTCAGGCT GCGCGGAGAC TCCCGTATCC AGGACGTTG
251 GTCCGCTCTG GCTCGGAGAG GATCTACTGT TCTCGGAGCC TCCTGGTGAC
301 GCGCTGCTCT CCGCTGCTGC ACCATAGGCC CGGAGTAGCG GCGTCCCGAC
351 AGCTTGGACC GGCAGGGGCT CGTGAATGT TTGTCAATG GATAAATGAC
401 CATGGCCGTG GTCTCGCGGG GAGGTGAGGA AACTGAAAGC CACCGAGGAA
451 AAGGGGGCGG CTCCTTAAGA AGTGCCCGGG TCACGTGTAC GTTTCAAAAG
501 AATGGCGTGA CTGAGTAGGG AGGGGACCGG GGAGACCCTC AGACCCCTGA
551 CTGTAAAGAG ATGAGGGGCG GTGAAGGGGA ACCCAGGAAA CTGAGTCTCTG
601 AAAGCAAGGA GGAACCTCCA GAATGAAGGC CCGCGAGACT CCTTCGTCC
651 TTTCGTCGAG CGGTTCCTTC ACCCGATCA AGTTCCTTCC CATTCCTCCA
701 TCTGGGGGAT CCTGAACGTG CACATCTCA GAGAAGCCCT CCGGGGTCT
751 CCAATTCTAG TTTATTGCCC CCTCTATCG ATCCCCCAGC GCGCTCATCG
801 GGGCTGTGGA CAAGGACAGG TTGAAGAGA GGATTCCTCT GATCGCGGAA
851 GGGCTGCAGG AATGGCACAG CCGCTTCCGA GGATGCCAAA GGAGCCCGGG
901 CAAAGGAAAG TGGCCGTGCC CGGGCCTGCC TACCACTAGA TCCCCACCCA
951 CCTATGACTG CTCAGTCCCG CTCTCTTACC ACACCCACCT TTCCGGGCCC
1001 AAGCCAGCGC ACCCGCTGGA CTCCCTGCCC AGTCCAAACT CCAAGGCTGG
1051 GCAAGGCACT GATCCACTGC TGGACAGACC CGGGGAGGCC TCTGGGTGAA
1101 CAGCAGCGTG TCCGCCGGCA GCGAACCGAG ACCAGCGAGC CGACCATGCG
1151 GCTGCACAGA CTTCGTGCGC GGCTGAGCGC GGTGGCCTGT GGGCTTCTGC
1201 TGCTTCTTGT CCGGGGCCAG GGCCAGGACT CAGCCAGTCC CATCCGGAAC
1251 ACACACACGG GGCAGGTGCT GGGGAGTCTT GTCCATGTGA AGGCGGCCAA
1301 TGCGGGGGTC CAAACCTTCC TGGGAATTCC ATTGCGAAG CCACCTCTAG
1351 GTCCGCTGCG ATTTCGACCC CCGTGAACCC CTGAATCTTG GAGTGGTGT
1401 AGGATGAGAA CCACCATATC GGCCATGTGT CTACAGGACC TCACCGCAGT
1451 GGACTCAGAG TTTCTTAGCC AGTTCAACAT GACCTTCCCT TCCGACTCCA
1501 TGCTGAGGGA CTGCTGTATC CTCAGCATCT ACACGCCGGC CCATAGCCAT
1551 GAAGGCTCTA ACCTGCCGGT GATGTTGTG ATCCACGGTG GTGCGCTTGT
1601 TTTTGGCATG GCTTCCTTGT ATGATGGTTC CATGCTGGCT GCCTTGAGAA
1651 ACGTGGTGGT GGTCAATATC CAGTACCGCC TGGGTGTCTT GGGCTTCTTC
1701 AGCACTGGAG ACAAGCACGC AACCGGCAAC TGGGGCTACC TGGACCAAGT
1751 GGCTGCACCTA CGCTGGGTCC AGCAGAAATAT CCGCCCTTTT GGAGCAACCT
1801 CTGACCTGTG CACCAATTTT GGCGAGTCTG CCGGTGGCAC GAGTGTGTCT
1851 TCGCTTGTGG TGTCGCCCAT ATCCCAAGGA CTCTTCCAGC GAGCATCAT
1901 GGAGAGTGGC GTGGCCCTCC TGCCCGGCTT CATTGCCAGC TCAGCTGATG
1951 TCATCTCCAC GGTGGTGGCC AACCTGTCTG CCTGTGACCA AGTGTACTCT
2001 GAGGCCCTGG TGGGCTGCCT GCGGGGCAAG AGTAAAGAGG AGATTCTTGC
2051 AATTAAACAAG CCTTCAAGA TGATCCCCGG AGTGGTGGAT GGGGTCTTCC
2101 TGCCCAAGGA CCCCCAGGAG CTGCTGGCTT CTGCGACTT TCAGCCTGTG
2151 CCTAGCATTG TTGGTGTCAA CAACAATGAA TTGGGCTGGC TCATCCCAA
2201 GGTCAAGAGG ATCATGATA CCGAGAGGA AATGGACAGA GAGGCTCTCC
2251 AGGCTGCTCT GCAGAAATG TTAACGCTGC TGATGTTGCC TCCTACATTT
2301 GGTGACCTGC TGAGGAGGGA GTACATTGGG GACAATGGGG ATCCCCAGAC
2351 CCTCCAAGCG CAGTTCCAGG AGATGATGGC GGAATCCATG TTTGTGATCC
2401 CTGCACTCCA AGTAGCACAT TTTCAGTGTT CCGGGGCCCC TGTGTACTTC
2451 TACGAGTTCC AGCATCAGCC CAGCTGGCTC AAGAATATCA GGGCACCGCA
2501 CATGAAGGCA GACCATGTTA AATTCACTGA GGAAGAGGAG CAGCTAAGCA
2551 GGAAGATGAT GAAGTACTGG GCAACTTTTG CCGAAATGTG GACCCCAAT
2601 GCGAGGGGTC TGCCACACTG GCGCTGTGTC GACAGGAGG AGCAATACCT
```

WO 01/12659

PCT/IB00/01496

```
2651 GCAGCTGAAC CTACAGCCTG CGGTGGGCGG GGCTCTGAAG GCCCAGAGC
2701 TCCAGTTCTG GAAGAAGGCG CTGCCCCAAA AGATCCAGGA GCTCGAGGAG
2751 CTTGAAGAGA GACACACAGA GCTGTAGCTC CCTGTCCCGG GGAGAGGGG
2801 CTGGGTTCCG TGACAGGCGA GGCTAGCCT GCTGTCCCA CACACACCA
2851 CTAAGGAGAA AGAAGTTGAT TCCTTCATAA AAAAAAAA
```

BLAST Results

Entry D50579 from database EMBL:
Homo sapiens mRNA for carboxylesterase, complete cds.
Score = 7197, P = 0.0e+00, identities = 1441/1443

Entry JC5408 from database PIR:
carboxylesterase (EC 3.1.1.1) - human
Score = 2808, P = 1.2e-291, identities = 542/559, positives = 543/559,
frame +3

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 954 bp to 2774 bp; peptide length: 607
Category: known protein
Classification: Metabolism
Prosite motifs: CARBOXYLESTERASE_B_1 (279-295)
CARBOXYLESTERASE_B_2 (185-196)

```
1 MTAQSRSPIT PTFPGPSQRT PLTPCPVQTP RLKALIHCV TDPGQPLGEQ
51 QRVRRQRTET SEPTNRLHRL RARLSAVACG LLLLIVRGQG QDSASFIRT
101 HTGQVLGSLV HVKANAGVQ TFLGIPFAKP PLGLPLRFAP EPPEWSGVR
151 DGTTHPAMCL QDLTAVESEF LSQFNMTFPS DSMSEDCLYL SIYTPAHSHE
201 GSNLPVMVWI HGGALVFGMA SLYDGSMLAA LENVVVVIIQ YRLGLVGFSS
251 TGDKHATGNW GYLDQVAALR WYQONIAHFG GNPDRVITFG ESAGGTSVSS
301 LUVSPISQGL FHGAIMESGV ALLPGLIASS ADVISTVVAN LSACDQVDSE
351 ALVGCLRGKS KEILAINKP FRMIPGVVDG VFLPRHPOEL LASADFPQVP
401 SIVGVNNEF GWLIPKVMRI YDTQKEMDRE ASQAALQKML TLLMLPPTG
451 DLLREYIGD NGDPQTLQAQ FQEMDADSMF VIFALQVAF QCSNAPVYFY
501 EFOHOPSLK NRPDPHMKAD HVKFTSEEEQ LSRKMMKYWA NFARNGNPNG
551 EGLPHMPLFD QEEQYLQNL QPAVGRALKA HRLQFWKKAL PQKIQELEEP
601 EERHTEL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n9, frame 3

PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human, N = 1, Score = 2808,
P = 1.9e-292

TREMBL:HSU60553_1 gene: "hCE-2"; product: "carboxylesterase"; Human
carboxylesterase (hCE-2) mRNA, complete cds., N = 1, Score = 2761, P =
1.8e-287

PIR:A34329 60K esterase (EC 3.1.1.-) isoform 2 - rabbit, N = 1, Score =
1985, P = 3.1e-205

TREMBL:D50580_1 product: "carboxylesterase precursor"; Rattus
norvegicus mRNA for carboxylesterase, partial cds., N = 1, Score =
1984, P = 4e-205

>PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human
Length = 559

HSPs:

Score = 2808 (421.3 bits), Expect = 1.9e-292, P = 1.9e-292

Identities = 542/559 (96%), Positives = 543/559 (97%)

Query: 65 MRLHRLRLRLSAVACGLLLLVRCQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLG 124
Sbjct: 1 MRLHRLRLRLSAVACGLLLLVRCQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLG 60

Query: 125 IPFAKPLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTPFSDSMS 184
Sbjct: 61 IPFAKPLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTPFSDSMS 120

Query: 185 EDCLYLSIYTPAHSHEGSLNPMVMIHGGALVFGMASLYDGSMLAALENVVVVIQYRLG 244
Sbjct: 121 EDCLYLSIYTPAHSHEGSLNPMVMIHGGALVFGMASLYDGSMLAALENVVVVIQYRLG 180

Query: 245 VLGFSTGDKHATGNMGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 304
Sbjct: 181 VLGFSTGDKHATGNMGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 240

Query: 305 PISQGLFHGAIMESGVALLPGLIASSADVISTVUANLSACDQVDSALVGCRLRGSKKEEI 364
Sbjct: 241 PISQGLFHGAIMESGVALLPGLIASSADVISTVUANLSACDQVDSALVGCRLRGSKKEEI 300

Query: 365 LAINKPFKMPGVVDGVFLPHRPOELLASADFPVPSIVGVNNNEFGWLI PKVMRIYDTQ 424
Sbjct: 301 LAINKPFKMPGVVDGVFLPHRPOELLASADFPVPSIVGVNNNEFGWLI PKVMRIYDTQ 360

Query: 425 KEMDREASQAALQKMLTLMLPPTFGDLLREEYIGDNGDPOTLQAQFOEMMADSMFVIPA 484
Sbjct: 361 KEMDREASQAALQKMLTLMLPPTFGDLLREEYIGDNGDPOTLQAQFOEMMADSMFVIPA 420

Query: 485 LQVAHFQCSRAPVYFYEFQHPQSWLKNIRPHMKADH-----VKFTEE 528
Sbjct: 421 LQVAHFQCSRAPVYFYEFQHPQSWLKNIRPHMKADH +KFTEE 480

Query: 529 EQLSRKMMKYWANFARNGNPNHGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 588
Sbjct: 481 EQLSRKMMKYWANFARNGNPNHGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 540

Query: 589 ALPQKIQELEEPEERHTEL 607
Sbjct: 541 ALPQKIQELEEPEERHTEL 559

Pedant information for DKF2phtes3_35n9, frame 3

Report for DKF2phtes3_35n9.3

(LENGTH) 607
(MW) 67051.20
(PI) 6.11
(HOMOL) PIR:JC5408 Carboxylesterase (EC 3.1.1.1) - human 0.0
(BLOCKS) BL01173A Lipolytic enzymes "G-D-X-G" family, histidine
(BLOCKS) BL00122G
(BLOCKS) BL00122F
(BLOCKS) BL00122S
(BLOCKS) BL00122D Carboxylesterases type-B serine proteins
(BLOCKS) BL00122C Carboxylesterases type-B serine proteins
(BLOCKS) BL00122B Carboxylesterases type-B serine proteins
(BLOCKS) BL00122A Carboxylesterases type-B serine proteins
(SCOP) d1akn 3.56.1.1.4 Bile-salt activated lipase [Bovine (Bos taurus 1e-158
(SCOP) d1thg 3.56.1.1.1 Acetylcholinesterase [Electric ray (Torped 1e-170
(SCOP) 3.1.1.13 Sterol esterase 1e-52
(EC) 3.1.1.7 Acetylcholinesterase 5e-74
(EC) 3.1.1.1 Carboxylesterase 0.0
(EC) 3.1.1.8 Cholinesterase 5e-68
(EC) 3.1.1.59 Juvenile-hormone esterase 1e-34
(EC) 3.1.1.3 Triacylglycerol lipase 3e-52
(PIRKW) duplication 2e-47
(PIRKW) homotetramer 3e-67
(PIRKW) transmembrane protein 9e-44
(PIRKW) microsome 1e-130
(PIRKW) pancreas 3e-52
(PIRKW) endoplasmic reticulum 1e-134
(PIRKW) homotrimer 1e-134
(PIRKW) phosphatidylinositol linkage 5e-74
(PIRKW) synapse 3e-73
(PIRKW) liver 1e-131
(PIRKW) heparin binding 3e-52

```
[PIRKW]      phosphoprotein 7e-25
[PIRKW]      glycoprotein 1e-134
[PIRKW]      thyroid hormone biosynthesis 2e-47
[PIRKW]      carboxylic ester hydrolase 0.0
[PIRKW]      monomer 2e-42
[PIRKW]      disulfide bond 2e-31
[PIRKW]      mammary gland 3e-52
[PIRKW]      alternative splicing 5e-74
[PIRKW]      iodine 2e-47
[PIRKW]      pyroglutamic acid 6e-39
[PIRKW]      hydrolase 1e-135
[PIRKW]      muscle 3e-73
[PIRKW]      thyroid gland 2e-47
[PIRKW]      membrane protein 3e-73
[PIRKW]      neurotransmitter degradation 3e-73
[PIRKW]      cholesterol 3e-52
[PIRKW]      homodimer 2e-47
[PIRKW]      nerve 3e-73
[SUPFAM]     cholinesterase 0.0
[SUPFAM]     triacylglycerol lipase 1e-32
[SUPFAM]     cholinesterase homology 0.0
[SUPFAM]     thyroglobulin 2e-47
[SUPFAM]     thyroglobulin type I repeat homology 2e-47
[SUPFAM]     juvenile-hormone esterase 2e-35
[SUPFAM]     probable lipolytic protein ybaC 1e-07
[PROSITE]    CARBOXYLESTERASE_B_2 1
[PROSITE]    CARBOXYLESTERASE_B_1 1
[PFAM]       Carboxylesterases
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY      3.95 %

SEQ      NTAQSRSPPTTPTFFGPSQRTPLTPCPVOTPRLGKALHCWTDPGQPLGEQQRVRRQRTET
SEG      .....XXXXXXXXX.....
lacj)-   .....

SEQ      SEPTMRLHRLRLRLSAVACGLLLLVRGQGQDSASPIRTHTGQVLGSLVHVKGANAGVQ
SEG      .....XXXXX.....
lacj)-   .....ETEEEECEEEEEETTEE--EE

SEQ      TFLGIPFAKPLGLRFPAPPEPPESWSGVRDGTTHPAMCLOLTAVESEFLSQFMTFFPS
SEG      EEEEECEETTTGGGTTTCCCECCCCCEEECCCCCCECCCCCTTTT--HHHHCCCCC
lacj)-   .....

SEQ      DMSSEDCYLSIYTPAHSHEGSLPVMVWIRGGALVFGMASLYDGSMLAALENVVVVIIQ
SEG      CCBTTTTCEEEEEET--TTTTTEEEEEECTTTTTCTTTTGCCHHHHHHHHCEEEEC
lacj)-   .....

SEQ      YRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQNIAPHFGNPRDVTIFGESAGGTSVSS
SEG      CCCCCGGCCCTTTTTCCHHHHHHHHHHHHHHCGGGGCEEEEEEEECCHHHHHHHH
lacj)-   .....

SEQ      LVVSPISQGLFHGAIMESGVALLPLGIASSADVISTVVANLSACQVDSALVGCLRGKS
SEG      HHHCGGGTTTCEEEEEETTTTTTTTBCCHHHHHHHHHHHHC-CCCCCHHHHHHHHHHCC
lacj)-   .....

SEQ      KEEILAINKPFKMI PGVVGVLPRHPQELLASADFQVPVSI VGNNEFGWLI PKVMRI
SEG      HHHHHHHHTCCCTTCBTTTTTTTTTTHHHHHHHTTCCCEEEEEETBTTHHHHHHTTTT
lacj)-   .....

SEQ      YDTQKENDREASQAALQKMLTLLMLPPTFGDLLREYIGDNGDPQLAQFQEMMADSMF
SEG      TTTCCCCCHHHHHHHHHHTTTCHHHHHHHHHHCTTTTTTHHHH--HHHHHHHHHHHH
lacj)-   .....

SEQ      VIPALQVAFQCSRAPVYFYEFOHQPSWLKNI RPPHMKADHVKT EEEELSRKMKYWA
SEG      HHHHHHHHHHHHCCCCCEEEECCECCGGGTBTTHHCGGGCCCHHHHHHHHHHHHH
lacj)-   .....

SEQ      NFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKKALPQKIQEELP
SEG      .....XXXXX.....
lacj)-   HHHHHCCCCC--CCCCBTTTBEEEECCCCCEETTTTHHHHHHHHHHHH.....

SEQ      EERHTEL
SEG      XXXXXX.
lacj)-   .....
```

Prosites for DKF2phtes3_35n9.3

PS00122 279->295 CARBOXYLESTERASE_B_1 PDOC00112
PS00941 185->196 CARBOXYLESTERASE_B_2 PDOC00112

Pfam for DKFZphtes3_35n9.3

HMM_NAME	Carboxylesterases
HMM	*MfMnWlImFLlWmItWii.WhegapspPdPyivdtnnCGkIRGmNedtD + +L+++ ++++++ ++Q+++P I T+ G+G++ +
Query	69 RLRARLSAVACGLLLLLVRGQGQDSASP---IRTTHT-GQVLGSLVHVK 113
HMM	NG..pYYvFLGIPYAEPVGNLRFKePQPYhePwtnVNNATnYPPMCMQW + + +FLGIP+A+PP+G LRF +P+P +E W++V++ T+ P MC+Q+
Query	114 GANAGVQTFLGIPFAKPLGLRFAPPEP-PESWSGVRDGTTHPAMCLQD 162
HMM	ndFGFWlEdmiewWneniP..eMSDECLYNVWTPHnckPNskLPVMVWI +++ +N++ P +MSDECLYL++TP+ + +S+LPVMVWI
Query	163 LTAV--ESEFLSQFNMTFPDSMSDECLYLSYTPAHSHEGSNLPVMVWI 210
HMM	HGGGFNFGSGhsYFliqYDgeylMeeNVIVvtINyRLGPFGLSTgDid HGG+++FG + ++YDG+ L++ ENV+VY I+YRLG++GF+STGD +
Query	211 HGGALVFGMA----SLYDGSMLAALENVVVVIQYRLGVLFSTGDKH 255
HMM	lPPHGNWGLWDQRMALQWQDNIAFGGDPNNITIFGESAGGNSVHlHML + GNWG++DQ++AL+WVO+NIA+FGG+P+++TIFGESAGG+SV+ ++
Query	256 AT--GNWGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLV 303
HMM	SYGGDNPPmKqLFHRAIMQSGsAmcPMvIQsnyNaRqRArFARimGCN S P + +LFH AIN+SG A+ P+I S++ + +A++ C+
Query	304 S-----PISQGLFHGAIMESGVALPLGLIASSA--DVISTVVARLSACD 345
HMM	rmDssEMiQCLRsKPwEELmdAtWnFwmMfYFPFLPMFFgPVIDGDDaPE + DS++++ CLR K+ EE+++++ +F + + +DG+
Query	346 QVDSEALVGCLRGKSKEILAINK---PFRMIPGV-----VDGV---- 381
HMM	aFIPDHPeemIKEGkFndVPMIIGYNnDEGIWFapMmMnfnWfdEdEwId F+P+HP+E++++ F VP I+G+NN E++W++P M + + +E++
Query	382 -FLPRHPQELLASADFQPVFSIVGVNNEFGWLIKVMRIYDT-QKEMDR 429
HMM	itNedWyeWMPYILFYrddmsNikDMDDYIdkvyEeYPgWMDrFPqESYW ++ + ++ M +L + + + D ++EEY+G+ + PQ
Query	430 EASQAALQKMLTLMLPPT-F-----GDLLREEYIGDNGD-PQTLQA 469
HMM	nLqDMFTDYLFWCpTrIhadnHRKhwgsPVYMYeFDHPPSFGYgQFFmWR ++Q+M+ D F++P + ++H++ +PVY+YEF+H PS +
Query	470 QFQENMADSNFVIP--ALQVAFQCSRAPVYTYEQHQPSW-----LKN 511
HMM	HWppWmgvdH* +PP+M++DH
Query	512 IRPPHMKADH 521
HMM	*tEEELissMRmHGYYWINFARhGNPNthnglCWpPyTsnEQYdMIME TEEE+ +S R HM+YW+NFA+GNPN+ GL+MP +++EQY++ +
Query	525 TEEEEQLS-RKMHKYWANFARNGPNGE--GLPHWPLFDQEEQYLQLNL 570
HMM	tIImIOMCrmrDPYCNFW* + +++++ + FW
Query	571 QPAVGRALKARR--LQFW 586

DKFZphtes3_35p17

group: testes derived

DKFZphtes3_35p17 encodes a novel 505 amino acid protein with weak similarity to
Proteins of the armadillo family.

Proteins of the armadillo family are involved in diverse cellular processes in higher
eukaryotes. Some of them, like armadillo, beta-catenin and plakoglobins have dual functions in
intercellular junctions and signalling cascades. Others, belonging to the importin-alpha-
subfamily are involved in NLS recognition and nuclear transport, while some members of the
armadillo family have as yet unknown functions. The novel protein shows similarity to S.
cerevisiae protein Yel013p (VAC8) and Danio rerio b-catenin, but contains no armadillo (arm)
repeats.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific
genes.

similarity to S.cerevisiae VAC8

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1966 bp

Poly A stretch at pos. 1956, polyadenylation signal at pos. 1935

```
1 AAGTCAATG TAAGATTGGT TCATTAAAA TACTGAAGGA AATCAGTCAT
51 AATCCTCAAA TCAGACAGAA TATTGTTGAC CTTGGGGGCT TACCAATTAT
101 GGTGAATATA CTTGATTCTC CACACAAGAG TCTAAAATGT TTGGCAGCCG
151 AGACTATCGC GAATGTTGCC AAGTTAAAA GAGCACGGCG GGTGGTGAGG
201 CAGCACGGGG GTATCACCAG ACTGGTTGCT CTACTAGACT GTGCACATGA
251 TTCCACAAAA CCTGCCCAAT CGAGTCTGTA TGAGGCCAGA GACGTGGAAG
301 TGCTCGCTG TGCGGCACGT GCCCTGTGCA GCTGCAGTAA GACTCATAGG
351 AATAAGAGAG CCATCCGCAA AGCTGGGGCG ATTCTCTGT TGCTGGGCT
401 GCTGAAGACT TCTCATGAAA ACATGCTAAT TCCAGTGGT GGGCATTGCG
451 AAGAGTGTGC ATCAGAGGAA AACTACCGGG CTGCAATCAA AGCAGAAAGG
501 ATCATTGAAA ACCTTGTCAG GAACCTAAAT AGTGAGATG AGCAGCTGCA
551 GGAGCACTGC GCCATGGCCA TTTACCAGTG TGCTGAAGAT AAGGAAACCC
601 GGGACCTCGT TAGGCTGCAC GGAGGACTTA AGCCCTTGGC CAGTCTACTC
651 AATAACACTG ACAATAAAGA GCGGTTAGCT GCTGTACAGG GGCCTATATG
701 GAATGTTTCC ATCAGCAAG AGATGTTTAC CAAGTTCCGG GAATACAAAG
751 CCATTGAAAC CTTGCTGGGA CTTCTAACAG ATCAGCCTGA AGAAGTACTT
801 GTGAATGTGG TTGGGGCCTT GGGAGAATGC TGCCAAGAAC GTGAAAACCG
851 AGTCATTGTC CGGAATGTG GTGGCATTCA ACCACTTGTG AACCTCCTTG
901 TTGGAATAAA CCAAGCTCTT CTTGTGAATG TTACAAAAGC AGTTGGTGCT
951 TGTGAGTAG AACCTGAAAG TATGATGATA ATTGATCGCT TAGATGGAGT
1001 TCGTTTGTGG TGCTCCCTGC TGAAAAATCC TCACCCAGAC GTGAAGGCCA
1051 GCGCAGCATG GGCACCTCTG CCATGCATCA AAAATSCAAA GATGCTGGG
1101 GAATGCTTCT GTTCTCTTGT TGGTGGTTTG GAACCTATTG TCAATTACTT
1151 GAAATCAGAT AACAAAGAAG TTCTGGCAAG TGTATGTGCT GCCATTACCA
1201 ACATAGCAAA AGATCAAGAA AATTTAGCTG TTATCAGAGA TCATGGAGTT
1251 GTTCTTTTAT TGTCCAAACT GGCAAAATACA AATAACAATA AATTGAGACA
1301 TCATCTAGCA GAAGCTATTT CACGTGTGCT TATGTGGGGC AGGAATAGAG
1351 TGGCCTTCGG TGAGCACAAA GCAGTGGGTC CACTAGTGCG TTATCTGAAA
1401 TCAATGACA CCAAGCTGCA TCGGGCGACA GCTCAGGCTT TGTACCACT
1451 CTGAGAGAC GCCGATAACT GCATGCATAT GCATGAGAT GTTGCAATGA
1501 AGCTTCTACT GGATATGCTT GGCTCCCTG ACCAGGATCT CCAGGAAGCT
1551 GCAGCTGGTT GTATATCCAA TATCCGCGAG CTGGCTCTTG CTACAGAGAA
1601 GGCAAGATAC ACTTGAAATT TAAATGGACA TTACAAGCTA TCAAATTCTA
1651 CATGACACAG GACATGTGAC TCCCATGGCC AGAAAGCCTA AATTGGGAAA
1701 CAGTTGTTAG CAAACCCCTT CAACUATCTA AATGAAACA CACAAATTGA
1751 AATGACACAG AATGTTTTTC ATCTGAAAT TGCATGGAGA CTTTGTGTTT
1801 TATTTAATGT TTTCGAGATA TGACATGTGA TAAGATGGAA AGCCAAATAA
1851 CTTGATATAA GTTCTAAGA ATATGAGAT ATAGTATAT GATGTATTTT
1901 TAGTCAAGTG ATGCTTTTGT ATTTGTGGCG ATTTTAATAA AGGATATGGC
1951 CTTCCCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

98413148:
Ye1013p (Vac8p), an armadillo repeat protein related to plakoglobin and importin alpha is associated with the yeast vacuole membrane.

98330438:
YEB3/VAC8 encodes a myristylated armadillo protein of the *Saccharomyces cerevisiae* vacuolar membrane that functions in vacuole fusion and inheritance.

98158703:
Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

Peptide information for frame 3

ORF from 99 bp to 1613 bp; peptide length: 505
Category: similarity to known protein
Classification: unset

```
1 MVNILOSPHK SLKCLAAETI ANVAKFKRAR RVVRQHGGIT KLVALIDCAH
51 DSTKPAQSSL YEARDVEVAR CGALALWSCS KSHTNKEAIR KAGGIPLLAR
101 LLKTSHEHML IPVVGTLQEC ASEENYRAAI KAERIENLV KNLSENELQ
151 QHCAMAIYQ CAEDKETRDL VRLHGGLKPL ASLLNWDKK ERLAAVTGAI
201 WKSISKENV TKFREYKAI E TLVGLLTDOO EEVLVNVVGA LGCECCQEREN
251 RVIVRKCGGI OPLVNLVGI NOALLNVVTK AVGACAVEPE SMHIIDRLDG
301 VRLWLSLLKN PHPDVKASAA WALCPCKIKNA KDAGEMVRSF VGGLELIVNL
351 LKSDNKEVLA SVCAAITNIA KDQENLAVIT DHGVVPLLSK LANTNNKLR
401 HHLAEAISRC CMWGRNRVAF GEHKAVAPLV RYLKSDNTNV HRATAQALYQ
451 LSEDADNCIT MHENGAVKLL LDMVGSPDQD LQEAAGCIS NIRRLALATE
501 KARIY
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p17, frame 3

PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 237, P = 7.8e-17

PIR:T00403 T13E15.9 protein - *Arabidopsis thaliana*, N = 1, Score = 215,
P = 4.9e-14

TREMBL:DR41081.1 product: "b-catenin"; Danio rerio b-catenin mRNA,
complete cds., N = 1, Score = 195, P = 5.8e-12

>PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*)
Length = 578

HSPs:

Score = 237 (35.6 bits), Expect = 7.8e-17, P = 7.8e-17
Identities = 106/401 (26%), Positives = 177/401 (44%)

```
Query: 92 AGGIFLLARLLKTSHEHMLIPVVGTLQECASEENYRAAIKAERIENLVKNLSENELQ 151
      +GG PL A +N+ + L E V + E ++E ++ L S++ O+Q
Sbjct: 45 SGG-PLKALTTLVYSDNLNQRSALAFAEITEKYVRQVSRE-VLEPIILLQSDPQIQ 102

Query: 152 EHCAMAIYQCAEDKETRDLVRLHGGLKPLASLLNNTDNKERLAAVTGAINWKSISKENVT 211
      A+ A + E + L+ GGL+PL + + DN E G I + +N
Sbjct: 103 VAACAALGNLAVNNENKLLIVEMGGLEPLINQMGM-DNVEVQCNAVGCITNLATRDONKH 161

Query: 212 KFREYKAIETLVGLLTDOPEEVLVNVVVGALGCECCQERENRIVRKCGGIOPLVNLLVGIN 271
      K A+ L L L + V N GAL ENR + G + LV+LL +
Sbjct: 162 KIATSGALIPLTCLKAKSHIRVQRNATGALLNTHSEENRKELVNAGAVPVLVSLLSSTD 221

Query: 272 QALLVNVTKAVGACAVEPESMHIIDRLDG--VRLWLSLLKNHPDVKASAAWALCPCKIKN 329
      + T A+ AV+ + + + + V L SL+ +P VK A AL +
```


Sbjct: 222 PDVQYCTTALSNIADENRKKLAQTEPRLVSKLVSLMDSPPSRVKQATLALRNLASD 281

Query: 330 AKDAGENVRSEFVGGLIIVNLLKSDNKE-VLASVCAAITNIAKDQENLAVITDHGVV-PL 387
E+VR+ GGL +V L++SD+ VLASV A I NI+ N +I D G + PL

Sbjct: 282 TSYQLEIVRA--GGIPLVKLIQSDSIPLVLASV-ACIRNISIHPLNEGLIVDAGFLKPL 338

Query: 388 LSKLANTNNKLRHHLAEASIRCCMMG-RNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQ 446
+ L ++ ++ H + +NR F E AV +V ++

Sbjct: 339 VRLLDYKDSEIQCIAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSV-QSEIS 397

Query: 447 ALYQLESDAD-NCITMHENGAVKLLDMVGSPPDQLQEAAGCISNI 492
A + + AD + + + E + L+ M S +Q++ AA ++N+

Sbjct: 398 ACFAILALADVSKLDLEANILDALIPMTFSQNEQVSGHAAALANL 444

Score = 213 (32.0 bits), Expect = 3.6e-14, P = 3.6e-14
Identities = 81/341 (23%), Positives = 163/341 (47%)

Query: 163 EDKETROLVRLHGGKPLASLLNNTD-NKERLAAVTGAIWKCSISKENVTKFREYKAIET 221
EDK+ D G LK L +L+ + + N+ R AA+ A I+++ V + + +E

Sbjct: 36 EDKDQDFYS-GGPLKALTTLVYSDNLNQRSAAALFA----EITEKYVRQVSR-EVLEP 89

Query: 222 LVGLLTQDPEEVLVNVVVGALGECCQERENRIVRKCGGIOPVLNVLVGIQALLVNVTKA 281
++ LL Q ++ V ALG EN++++ + GG++PL+N ++G N + N

Sbjct: 90 ILILLOSQDPQIQVAAALGNLAVNNENKLLIVEMGGLEPLNQMGDNVEVQCNVAGC 149

Query: 282 VGACAVEPESMHIIIDRLDGVRLLSLLKNPHDPVKASAAWALCPCKNAKDAGEMVRSFV 341
+ A ++ I + L L K+ H V+ +A AL + + + E+V +

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIVQRNATGALLNMTHSEENRKLVA-- 207

Query: 342 GGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVI--TDHGVVPLLSKLANTNNKXL 399
G + ++V+LL S + +V A++NIA D+ N + T+ +V L L ++ + + +

Sbjct: 208 GAVPVLVSLSSSTDPVQYCTTALSNIADENRKKLAQTEPRLVSKLVSLMDSPPSRV 267

Query: 400 RHHLAEASIRCCMMGRNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQALYQLESDADNCI 459
+ A + + + + + LV+ ++S+ + A+ + +S N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query: 460 TMHENGAVKLLDMVGSPPDQLQEAAGCISNIRRLALATEKAR 503
+ + G +K L+ ++ D + E +S +R LA ++EK R

Sbjct: 328 LIVDAGFLKPLVRLLDYKDSE--EIQCHAVSTLRNLAASSEKNR 369

Score = 180 (27.0 bits), Expect = 1.6e-10, P = 1.6e-10
Identities = 80/346 (23%), Positives = 142/346 (41%)

Query: 145 SENEQLQEHCAIAIYQCAEDKETDRLVRLHGGKPLASLLNNTDNKERLAAVTGAIWKCS 204
S+N LQ A+A + E K R + R L+P+ LL + D + ++AA A+ +

Sbjct: 58 SDNLNQRSAAALAFABEITE-KYVRQVSR--EVLEPILILLOSQDPQIQVAAAL-ALGNLA 113

Query: 205 ISKENVTKFREYKAIETLVGLLTQDPEEVLVNVVVGALGECCQERENRIVRKCGGIOPLV 264
++ EN E +E L+ + + EV N VG + +N+ + G + PL

Sbjct: 114 VNNENKLLIVEMGGLEPLNQMGDNVEVQCNVAGCITNLATRDNDKHKIATSGALIPLT 173

Query: 265 NLLVGINQALLVNVTKAVGACAVEPESMHIIIDRLDGVRLLSLLKNPHDPVKASAAWALC 324
L + + N T A+ E+ + V+L SLL + PDV+ AL

Sbjct: 174 KLAKSKHIVQRNATGALLNMTHSEENRKLVAAGAVPVLVSLSSSTDPVQYCTTALS 233

Query: 325 PCIKNAKDAGEMVRSFVGGLIIVNLLKSDNKEVLASVCAAITNIAKDQENLAVITDHGV 384
+ + + + + + + +V+L+ S + V A+ N+A D I G

Sbjct: 234 NIAVDEANRKKLAQTEPRLVSKLVSLMDSPPSRVKQATLALRNLASDTSYQLEIVRAGG 293

Query: 385 VPLLSKLANTNNKLRHHLAEASIRCCMMGRNRVAFGEHKAVAPLVRYLKSNDTNVHRAT 444
+P L KL +++ L I + N + + + PLVR L D+ +

Sbjct: 294 LPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDSEIQC 353

Query: 445 A-QALYQLESDAD-NCITMHENGAVKLLDMVGSPPDQLQEAAGCIS 490
A L L+ ++ N E+GAV+ ++ +Q + C +

Sbjct: 354 AVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFA 401

Score = 155 (23.3 bits), Expect = 8.6e-08, P = 8.6e-08
Identities = 88/401 (21%), Positives = 175/401 (43%)

Query: 60 LYEAR--VEVARCGALALWCSKSHNTKEAIRKAGGI-PLLARLLKTSHENMLIPVVG 116
L +++D ++VA C AL + + + NK I + GG+ PL+ ++ + E + VG

Sbjct: 93 LLQSODPQIQVAAALG--NLAVNNENKLLIVEMGGLEPLNQMGDNVE-VQCNVAGC 149

Query: 117 LQECASEENVRAIKAEIRIENLVKNLSENEQLQEHCAIAIYQCAEDKETR-DLVRLHG 175
+ A + + + I + L K S++ ++Q + A+ +ER +LV G

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIVQRNATGALLNMTHSEENRKLVA-- 208

Query: 176 GLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFR--EYKAIETLVGLLTQDPEEV 233
+ L SLL++TD + T A+ ++ + N K E + + LV L+ V

Sbjct: 209 AVPVLSLLSSTDPDVQYCTT-ALSNIADVEANRKKLAQTEPRVSKVSLMDSPPSSRV 267

Query: 234 LVNVVVGALGECQERENRVRKCGGIQPLVNLVGINOALLVNVTKAVGACAVEPESMM 293
AL + + + + GG+ LV L+ + L+ + + + P +

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query: 294 IIDRLDGVRLLSLLK-NPHPDVKASAAWALCPCIKNA-KDAGEMVRSFVGGLLITVNL 351
+I + + L LL + + + A L + + K+ E S G +E+ L

Sbjct: 328 LIVDAGFLKPLVRLLDYKDEEIQCHAVSTLRNLAASSEKNRKEFFES--GAVERCKELA 385

Query: 352 KSDNKEVLA--SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEAISR 409
V + S C AI +A D L + + + + L + + N + + A A + +

Sbjct: 386 LDSPVSVQSEISACFAILALA-DVSKLDL-EANILDALIPMTFSQNEVSGNAAAALAN 443

Query: 410 CCHMGRRNVAPE-----HKAVAP-LVRYLKSNDTNVHRATAQALYQLE 453
C N E + + + L+R+LKS+ + QL E

Sbjct: 444 LCSRNNYTKIIEAWDRPNEGIRGFLIRFLKSDYATFEHIALWTILQLE 493

Score = 139 (20.9 bits), Expect = 5.0e-06, P = 5.0e-06
Identities = 80/329 (24%), Positives = 142/329 (43%)

Query: 37 GGITKLVALLOCAHD-STKPAQ---SSLYEARDVEVARCAGALWSCSKSHTNKEAIRKA 92
C IT L D H +T A L + + + + V R AL + + S N + + A

Sbjct: 148 GCITHLATRDNDKHKIATSGALIPLTKLAKSKHVRQVORNTGALLNMTHSEENRKELVNA 207

Query: 93 GGIPLLARLLKTSHEENMLIPVVGTLQECASEE-NYRAATKAE-RIENLVKNINSENEQL 150
G +P+L LL + + + L A +E N + + E R+ + LV + +S + + +

Sbjct: 208 GAVPVLVLSLLSSTDPDVQYCTTALSNIADVEANRKKLAQTEPRVSKVSLMDSPPSSRV 267

Query: 151 QEHCAIYQCAEDKETR-DLVRHGGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKEN 209
+ +A+ A D + + +VR GGL L L+ + D+ + A I SI N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRA-GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLN 325

Query: 210 VTKFREYKAIEITLVGLLT-DQPEEVLNVVVGALGECQERE-NRIVRRCGGIQLVNL 267
+ + + + + LV LL EE+ + V L E R + G + + L

Sbjct: 326 EGLIVDAGFLKPLVRLLDYKDEEIQCHAVSTLRNLAASSEKNRKEFFESGAVERCKELA 385

Query: 268 VG--INQALLVNVTKAVGACA-VEPESMMIIDRLDGVRLLSLLKNPHPDVKASAAWA-L 323
+ + + + + A A A V + + + LD + + + +H A+AA A L

Sbjct: 386 LDSPVSVQSEISACFAILALADVSKLDLEANILDAL-IPMTFSQNEVSGNAAAALANL 444

Query: 324 CPCIKN-AKDAGEMVRSFVGGLLITVNLKSD 354
C + N K R C + + LKSD

Sbjct: 445 CSRNNYTKIIEAWDRPNEGIRGFLIRFLKSD 476

Score = 136 (20.4 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 72/304 (23%), Positives = 133/304 (43%)

Query: 58 SSLYEARDVEVARCAGALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEENMLIPVVGTL 117
+ L + + + + V R AL + + S N + + + AG +P+L LL + + + L

Sbjct: 173 TKLAKSKHVRQVORNTGALLNMTHSEENRKELVNAGAVPVLVLSSTDPDVQYCTTAL 232

Query: 118 QECASEE-NYRAATKAE-RIENLVKNINSENEQLQEHCAIYQCAEDKETR-DLVRH 174
A +E N + + E R+ + LV + +S + + + +A+ A D + +VR

Sbjct: 233 SNIADVEANRKKLAQTEPRVSKVSLMDSPPSSRVKCQATLALRNLASDTSYQLEIVRA- 291

Query: 175 GGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIEITLVGLLT-DQPEEV 233
GGL L L+ + D+ + A I SI N + + + LV LL EE+

Sbjct: 292 GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDEEI 350

Query: 234 LVNVVVGALGECQERE-NRIVRRCGGIQLVNLVVG--INQALLVNVTKAVGACA-VEP 289
+ V L E R + G + + L + + + + A+ A A V

Sbjct: 351 QCHAVSTLRNLAASSEKNRKEFFESGAVERCKELALDSPVSVQSEISACFAILADVSK 410

Query: 290 ESMNIIDRLDGVRLLSLLKNPHPDVKASAAWA-LCPCIKN-AKDAGEMVRSFVGGLLI 347
+ + + LD + + + +N A+AA A LC + N K R G +

Sbjct: 411 LDLEANILDAL-IPMTFSQNEVSGNAAAALANLCSRNNYTKIIEAWDRPNEGIRGFL 469

Query: 348 VNLKSD 354
+ LKSD

Sbjct: 470 IRFLKSD 476

Score = 114 (17.1 bits), Expect = 2.7e-03, P = 2.7e-03
Identities = 71/335 (21%), Positives = 132/335 (39%)

Query: 1 MVNILDSPHKSLEKLAETIANVAKFKRARRVVRHGGITKLVALLOCAHDSTKPAQSSL 60
+ + S H + + A + N+ + R+ + G + LV+LL ST P

Sbjct: 172 LTKLAKSKHVRQVORNTGALLNMTHSEENRKELVNAGAVPVLVLSLLS-----STDP----- 222

Query: 61 YEARDVEVARCAGALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEENMLIPVVGTLQEC 120
DV+ AL+ + + + K A + + L L+ + + L+

Sbjct: 223 ----DVQYYCTTALSNIADVEANRKKLAQTEPRLVSKLVSLMDSPPSRVKQATLALRNL 278

Query: 121 ASEENYRAAIKAERIIENLVKNLSENEQLQEHCAAIYQCAEDKETRDLVRLHGGLKPL 180
AS+ +Y+ I + +LVK + S++ L I + L+ G LKPL

Sbjct: 279 ASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPL 338

Query: 181 ASLLNNTONKERLAAVTGAIWKCSISKE-NVTKFREYKAIETLVGLLTDQPEEVLNVVVG 239
LL+ D++E + + S E N +F E A+E L D P V +

Sbjct: 339 VRLDYKDSSEIQCHAVSTLRNLAASSEKNRKEFFESGAVECKELALDPSVVSQSEISA 398

Query: 240 ALGECCQERENRIVIRKCGGIQPLVNLVGINQALLVNVTKAVG-ACAVEPESMMIIDL 298
+++ + + + L+ + NQ + N A+ C+ II+

Sbjct: 399 CFALALADVSKLDLLEANILDALIPMTFSQNEVSGNAAALANLCSRNNYTKIIEAW 458

Query: 299 D----GVR-LLWSLLKNPDPVKASAAWALCPCKNAKDAGE 335
D G+R L LK+ + + A W + +++ D E

Sbjct: 459 DRPNEGIRGFLIRFLKSDYATFEHIALWTILQLLESHNDKVE 500

Score = 106 (15.9 bits), Expect = 2.0e-02, P = 2.0e-02
Identities = 49/204 (24%), Positives = 89/204 (43%)

Query: 65 DVEVARCGALA-LWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEHMLIPVVGTLQECA-S 122
+VEV +C A+ + + + NK I +G + L +L K+ H + G L S

Sbjct: 139 NVEV-QCNAVGCITNLATRDNNKKIATSGALIPLTKLAKSKHIVQRNATGALLNMTHS 197

Query: 123 EENYRAAIKAERIIENLVKNLSENEQLQEHCAAIYQCAEDKETRD-LVRLHGGL-KPL 180
EEN + + A + + LV L+S + +Q +C A+ A D+ R L + + L L

Sbjct: 198 EENRKELVNAGAV-PVLVSLSSDTPDVQYYCTTALSNIADVEANRKKLAQTEPRLVSKL 256

Query: 181 ASLLNNTONKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLNVVVG 240
SL+++ ++ + A T A+ + + + LV L+ +++ V

Sbjct: 257 VSLMDSPPSRVKQA-TLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVAC 315

Query: 241 LGECCQERENRIVIRKCGGIQPLVNL 267
+ + N + + G + +PLV LL

Sbjct: 316 IRNISIHLNEGLIVDAGFLKPLVRL 342

Pedant information for DKFZphtes3_35p17, frame 3

Report for DKFZphtes3_35p17.3

[LENGTH] 505
[MW] 55224.34
[PI] 8.43
[HOMOL] PIR:550446 VAC8 protein - yeast (Saccharomyces cerevisiae) 2e-16
[FUNCAT] 30.25 vacuolar and lysosomal organization [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YEL013w] 8e-18
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YNL189w] 3e-06
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL189w] 3e-06
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL189w] 3e-06
[BLOCKS] BL01265C
[BLOCKS] BL00242A Integrins alpha chain proteins
[SCOP] d3bct_1.91.1.1.1 beta-Catenin [Mouse (Mus musculus) 7e-18
[PIRKM] cytosol 3e-11
[PIRKM] apoptosis 3e-11
[PIRKM] carcinogenesis 3e-11
[PIRKM] cell adhesion 3e-11
[PIRKM] cytoskeleton 3e-12
[SUPFAM] pendulin 1e-07
[KW] All_Alpha
[KW] 3D
[KW] LOW_COMPLEXITY 2.38 %

SEQ MVNILDSPHKSCLKLAAETIANVAKFRARRVVRQHGGITKLVALLOCAHDSTKPAQSSSL
SEGXXXXXXXXXXXX.....RH
2bct-.....RH

SEQ YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHEHMLIPVVGTLQEC
SEGXXXXXXXXXXXX.....RH
2bct-.....RH

SEQ ASEENYRAAIKAERIIENLVKNLSENEQLQEHCAAIYQCAEDKETRDLVRLHGGLKPL
SEGXXXXXXXXXXXX.....RH
2bct-.....RH

WO 01/12659

PCT/IB00/01496

```
SEQ ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTQQPEEVLNVVGA
SEG .....
2bct- HHHHH-HCCCCHHHHHHHHHHHHHHCCCCHHHHHHHHHCHHHHHHTTTTCCHHHHHHHHHH

SEQ LGECQCQERENRIVRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESHMIIDRLDG
SEG .....
2bct- H-----HHHHHCCCCCTTTTHHHHHHHHHHHHCTTTTHHHHHHHHTTTNHHHHHH-HHCH

SEQ VRLWSSLKNPHPDVKASAAWALCPCKNAKDAGENVRSFVGGLELIVNLLKSDNKEVLA
SEG .....
2bct- HHHHHHHHTTTTHHHHHHHHHHHHHHHCCCCCH-HHHHHHHHHHHHHHCTTTTHHHH

SEQ SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEISRCCMWGRNRVAF
SEG .....
2bct- HHHHHHHHHHHC GGHHHHHHHHHHHHHHHHHHHHHTTTTCCHHHHHHHHHHHCHHHHH

SEQ GENKAVAPLVRYLKSNDTNVHRATAQALYQLSEDAENCITMHENGAVKLLLDWVGS PDOD
SEG .....
2bct- HTTTTHHHHHHHHCCCCHHHHHHHHHHHHHHHTTTTHHHHHHHHCCCCHHHHHNTTTTTHH

SEQ LQEAAGCISNIRRLALATEKARYT
SEG .....
2bct- HHHHHHHHH.....
```

(No Prosite data available for DKFZphtes3_35p17.3)

(No Pfam data available for DKFZphtes3_35p17.3)

DKFZphtes3_35p22

group: cell cycle

DKFZphtes3_35p22 encodes a novel 549 amino acid protein, with similarity to oncogene 1 (tre-2 locus).

The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control.

The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

strong similarity to oncogene 1 (tre-2 locus)

membrane regions: 1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: map="17"

Insert length: 2072 bp

Poly A stretch at pos. 2062, polyadenylation signal at pos. 2039

```
1 GTTACACACA GGCAGTGGTA TCTGTGAGCA GCTCTGTGGA CTCAAAGGTT
51 TTCTCCCTGA GAGGCATGAC CCAGGCCAGC TGATTCATCA GAATCAGGAT
101 GGACGTGGTA GAGGTCCGGG GCAGTTGGTG GGCACAAGAG CGAGAGGACA
151 TCATTATGAA ATACGAAAAG GGACACCGAG CTGGGCTGCC AGAGGACAAAG
201 GGGCCTAAGC CTTTTCGAAG CTACACAAAC AACCTGCATC ATTTGGGGAT
251 TGTACATGAG ACGGAGCTGC CTCCTCTGAC TGGCGGGAG GCGAAGCAAA
301 TTGGCGGGGA GATCAGCCGA AAGAGCAAGT GGGTGGATAT GCTGGGAGAC
351 TGGGAGAAAT ACAAAAGCAG CAGAAAGCTC ATAGATCGAG CGTACAAGGG
401 AATGCCCATG AACATCCGGG GCCCGATGTG GTCAGTCCCT CTGAACACTG
451 AGGAAATGAA GTTGA AAAAC CCCGGAAGAT ACCAGATCAT GAAGGAGAG
501 GGCAGAAGT CATCTGAGCA CATCCAGCGC ATGACCGGG ACCTAAGCGG
551 GACATTAGG AGCATATAT TCTTCAGGCA TCGATACGGA ACCAAGCAGC
601 GGGAACTACT CCACATCCTC CTGGCATATG AGGAGTACAA CCCGGAGGTG
651 GGCTACTGCA GGGACCTGAG CCACATCGCC GCCTTGTTC TCTCTATCT
701 TCTGAGGAG GATGCATTCT GGGCACTGGT GCAGCTGCTG GCCAGTGAGA
751 GGCCTCCCT GCAGGGATT TCCAGCCCAA ATGGCGGGAC CTTCAAGGGG
801 CTCCAGAGCC AACAGGACCA TGTGTTAGCC ACCTGCAAC CCAAGACCAT
851 GGGCATCAG GACAAGAAAG ATCTATGTGG GCAGTGTTC CCCTTAGGCT
901 GCCTCATCCG GATATTGATT GACGGGATCT CTCCTGGGCT CACCTGCGC
951 CTGTGGGACG TGTATCTGTT AGAAGCGGAA CAGGCGCTGA TGCCGATAAC
1001 AAGAAATGCC TTTAAGGTTT AGCAGAAGCG CCTCAGGAAG ACCTCCAGGT
1051 GTGGCCCGTG GGCACGTTTT TGCAACCGGT TCGTTGATAC CTGGGCCAGG
1101 GATGAGGACA CTGTGCTCAA GCATCTTAGG GCCTCTATGA AGAATTAAC
1151 AAGAAAGAG GGGGACCTGC CACCCCGCAG CAAACCCGAG CAGGGGTGCT
1201 GGGCATCCAG GCCTGTGCCG GCTTCAGTGG GCGGGAAGAC CCTCTGCAAG
1251 GGGGACAGGC AGGCCCTTCC AGGCCACCCA GCGCGGTTCC CGCGGCCCAT
1301 TTGGTCAGCT TCCCGGCCAC GGGCACCTCG TTCTTCCACA CCCTGTCCCTG
1351 GTGGGCTGT CCGGGAAGAC ACCTACCTTG TGGGCACTCA GGGTGTGCC
1401 AGCCCGGCC TGGCTCAGGG AGGACTCAG GGTTCCTGGA GATTCTGCA
1451 GTGAACCTC ATGCCCGCC TCCCAAGGGA CTTGGACGTA GAGGGCCCTT
1501 GGTTCGCCCA TTATGATTTC AGACAGAGCT GCTGGGTCCG TGCCATATCC
1551 CAGGAGGACC AGCTGGCCCC CTGCTGGCAG GCTGAACACC CTGCGGAGCG
1601 GGTGAGATCG GCTTTCGCTG CACCCAGCAC TGATTCCGAC CAGGGCACCC
1651 CTTTCAAGAG TAGGAGCGAA CAGCAGTGTG CTCCCACTC AGGGCCTTGG
1701 CTCTGGGCC TCCACTTGGG AAGTTCTCAG TTCCCTCAG GCTCTAGAA
1751 GCATCTGGC CAGGCTCAT GCGTGATAA TTCCCTAGG CTTAACAACC
1801 CAAGCAAGCT TCGCATCCTC GTTTTATTTT TGGTTAAACT TATGAAAATG
1851 TATTAAGAAA GAGTGCAGCT CGAGAGAGAT TCAGAGATGG AACACACGAG
1901 ACCCCAGATC ACAAGGCCAA CCATGCCAG CCCCTCCAG CACCCACGAG
1951 CCCACGACCA TCGTTCTGAA TTCTGACGAC ACCGTGAGCC TGCCCTTGTG
2001 CTCRAAATC ATGGAAGGAT AACCACTTC ATGTTTTGAA ATAAATGTTT
2051 CCGTTGAAA TGAAAAAAA AA
```

BLAST Results

Entry AC003976 from database EMBL:
Homo sapiens chromosome 17, clone hCIT.91 J.4, complete sequence.
Score = 4385, P = 0.0e+00, identities = 8817886

14 exons

Entry HSG19723 from database EMBL:
human STS A001W35.
Score = 850, P = 1.9e-32, identities = 170/170

Medline entries

92228503:
A novel transcriptional unit of the tre oncogene widely
expressed in human cancer cells.

94067315:
The yeast DOA4 gene encodes a deubiquitinating enzyme
related to a product of the human tre-2 oncogene.

95176708:
UBP5 encodes a putative yeast ubiquitin-specific protease
that is related to the human Tre-2 oncogene product.

Peptide information for frame 3

ORF from 99 bp to 1745 bp; peptide length: 549
Category: strong similarity to known protein

```
1 MDVVEVAGSW WAQEREDIIM KYEGKHRAGL PEDKGPFR SYNNVVDHLG
51 IVHETELPPL TAREAKQIRR EISRSKWFV MLGOMEKYKS SRKLIDRAYK
101 GMPNMIIRGPH WSVLLNTEEM KLNKPGRYQI MKEKGKSSSE HIQRIORVVS
151 GTLRKHIFFR DRYGTQKREL LHILLAYEY NPEVGYCROD SHIAALFLLY
201 LPEEDAFWAL VOLLASERHS LQGFHSPNGG TVQGLQDQOE HVVATSQPKT
251 MGHQDKKDLG GOCSPGLCLI RILIDGSLG LTLRLWDVYL VEGEQALMPI
301 TRIAFKVOQR RLTKTSRCGP WAFPCNRFVD TWARDDETVL KALRASMKKL
351 TRAKGDLPPF AKFEQSSAS RPFVPSRGGK TLCKGQKQAP PQPPARFRFP
401 IWSASPPRAP RSTETPCGGA VREDTYPVGT QCVPSPALAQ GPGQSHRFL
451 QWNSMPRLPT DLDVEGPFWR HYDFRQSCWV RAISQEDOLA PCWQAEHPAE
501 RVRSAFAAPS TSDQGTFFR ARDEQQAFT SGPCLCGLHL ESSQFPFPGF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p22, frame 3

PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human, N = 1, Score =
2181, P = 5.5e-226

PIR:S57867 oncogene 1 - human, N = 1, Score = 1536, P = 1.2e-157

>PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human
Length = 786

HSPs:

Score = 2181 (327.2 bits), Expect = 5.5e-226, P = 5.5e-226
Identities = 405/500 (81%), Positives = 440/500 (88%)

```
Query: 1 MDVVEVAGSWWAQEREDIIMKYEGKHRAGLPEDKGPFRSYNNVVDHLGIVHETELPPL 60
MD+VE A S AQER+DI+MKY+KGHRAGLPEDKGP+P N+++D GI+HETELPP+
Sbjct: 1 MDNVENADSLQAQERKIDILMKYDKGHRAGLPEDKGPFPV-GINSSIDRFGILHETELPPV 59

Query: 61 TAREAKQIRREISRSKWFVMLGOMEKYKSSRKLIDRAYKGMNIRGPMMSVLLNTEEM 120
TAREAK+IRRE++R SKW++MLG+WE YK S KLIDR YKG+PMNIRGP+MSVLLN +E+
Sbjct: 60 TAREAKKIRREMTRTSKWMEMLGEWETKYKSSKLIDRVYKGI PMNIRGPMVSVLLNIQEI 119

Query: 121 KLENPGRYQIMKEGKSSSEHIQRIORDVSGTLRKHHIFFRDRYGTQKRELLHILLAYEY 180
KLENPGRYQIMKE+GK+SSEHI ID DV TLR H+FRDRYG KQREL +ILLAY EY
Sbjct: 120 KLENPGRYQIMKEGKSSSEHIHIDLVRTTLRNHVFRDRYGAKQRELFYILLAYEY 179

Query: 181 NPEVGYCRDLSHIAALFLLYLPEEDAFWALVOLLASERHS LQGFHSPNGGTVQGLQDQOE 240
NPEVGYCRDLSHI ALFLLYLPEEDAFWALVOLLASERHS L GFHSPNGGTVQGLQDQOE
Sbjct: 180 NPEVGYCRDLSHITALFLLYLPEEDAFWALVOLLASERHS LQGFHSPNGGTVQGLQDQOE 239
```

Query:	241	HVVATSPQKMHQOQKGLCCQGSGLGLRILR1DGTSLGLTLRLWDVYLVVEGGALPPI	300
Sbjct:	240	HVV SPKTM HOKAR + LCGOC + LGLLR + LDTGLSLTLRLWDVYLVVEGG LMP1	299
Query:	242	HVVFKSPQKMHQOQKGLCCQGSGLGLRILR1DGTSLGLTLRLWDVYLVVEGGALPPI	300
Sbjct:	240	HVVFKSPQKMHQOQKGLCCQGSGLGLRILR1DGTSLGLTLRLWDVYLVVEGGALPPI	299
Query:	243	TRIAFIQVQKRLKTSRCGPFARFNCNFTDWARDVETLHLKASHMKSTRKRGGLP	360
Sbjct:	300	T I A KVOQRL KTSRCG HAR + DFWA + DDTGLSLTLRLWDVYLVVEGG +GLPFP	359
Query:	244	TSIALVQKRLKTSRCGLMARLNRQVFTWMANNDDTVLHKRASTKTRKGGLP	360
Sbjct:	300	TSIALVQKRLKTSRCGLMARLNRQVFTWMANNDDTVLHKRASTKTRKGGLP	359
Query:	361	AKRQEGSSASRPVPSARSGKTKLCKDGRQAPGPCPPAREFP1PWSASPPRAPRSTPCPGA	420
Sbjct:	360	AKR EGS A RPVASRSGKTKLCKGR QAPGPCPPA RPI SASP A R STPCPGA	419
Query:	362	AKRQEGSLAPRPVPSARSGKTKLCKGYRQAPGPCPPAQFQAPRSPWASSRSTPCPGA	420
Sbjct:	360	AKRQEGSLAPRPVPSARSGKTKLCKGYRQAPGPCPPAQFQAPRSPWASSRSTPCPGA	419
Query:	421	VREDTYPVTGQVPSALAAQGGQGSWRFLEWNSMRLPTDLOVEGFWPHFYDFSCWCW	480
Sbjct:	420	VREDTYPVTGQVPSALAAQGGQGSWRFLEWNSMRLPTDLOVEGFWPHFYDFSCWCW	479
Query:	481	RAISQEDGLAPCQWAEHCE 500	
Sbjct:	480	RAISQEDGLAPCQWAEHCE 500	

Pedant information for DKFZphtes3_35p22, frame 3

Report for DKF2phtes3 35p22.3

```

[LENGTH]          549
[MM]               62159.16
[PI]               9.23
[HOMOL]            PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human 0.0
[FUNCAT]           11.01 stress response [S. cerevisiae, YGR100w] 2e-16
[FUNCAT]           04.05.01.04 transcriptional control [S. cerevisiae YGR100w] 2e-16
[FUNCAT]           99 unclassified proteins [S. cerevisiae, YNL293w] 3e-15
[PIKMW]            transmembrane protein 6e-14
[PROSITE]          MYRISTYL 6
[PROSITE]          AMIDATION 1
[PROSITE]          CAMP_PHOSPHO_SITE 3
[PROSITE]          CK2_PHOSPHO_SITE 4
[PROSITE]          TYR_PHOSPHO_SITE 2
[PROSITE]          PKC_PHOSPHO_SITE 10
[SW]               TRANSMEMBRANE 1
[SW]               LOW_COMPLEXITY 5.28 %

```

[illegible]

WO 01/12659

PCT/IB00/01496

```
SEQ VREDTYPVGTQGVPSPALAQGGPQGSWRFLQNSMRPLPTDLDVEGPFWRHYDFRQSCWV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ RAISQEDQLAPCQAEHPAERVRSFAAAPSTDSQGTFFRAREQQCAPTSGPCLGLHL
SEG .....
PRD cchhhhhhhhhhhhhcchhhhhhhccccccccccccchhhhhcccccccccccccccc
MEM .....

SEQ ESSQFPFGF
SEG .....
PRD ccccccccc
MEM .....
```

Prosite for DKF2phtes3_35p22.3

PS00004	136->140	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	310->314	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	348->352	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	73->76	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	152->155	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	446->449	PKC_PHOSPHO_SITE	PDOC00005
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00007	93->100	TYR_PHOSPHO_SITE	PDOC00007
PS00007	92->100	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	101->107	MYRISTYL	PDOC00008
PS00008	230->236	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	441->447	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_35p22.3)

DKFZphtes3_4b4

group: testes derived

DKFZphtes3_4b4 encodes a novel 497 amino acid protein similar to SCP proteins and a human trypsin inhibitor.

The novel protein contains an extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2, predicted by Prosite and Pfam. This domain is found in a variety of extracellular proteins from eukaryotes that have been found to be evolutionary related. The exact function of these proteins is not yet known. In addition, the protein is similar to a human trypsin inhibitor.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes or as a new protease inhibitor.

strong similarity to trypsin inhibitor

might be a new protease inhibitor?

Sequenced by AGOWA

Locus: /map="333.4 cR from top of Chr16 linkage group"

Insert length: 4574 bp

Poly A stretch at pos. 4551, polyadenylation signal at pos. 4539

```
1 GCGCGCTGCT CCCATTGAGC TGCTGCTCG CTGTGCCCGC TGTGCTGCT
51 GTGCGCCGCG GTGCGCCGCT GCTACCGGGT CTGTGGGAGC CGGAGAGCG
101 CAGCGAGCTG GTGATTGGAG CCTGGGGAG AGCTCAAGCG CCGAGCTCTG
151 CCGAGGAGAG CCAAGGCTGCC CCGTGAGTCC CATAGTTGCT GCAGGAGTGG
201 AGCCTAGTGC TGCCTCTGCG CTGGGTGCTA CCGCTGGGGC CTGCTGTTC
251 TGGTCTGCGG ATCCCAAGGC TACTCTCTGC CCAAGCTCAC TCTCTTAGAG
301 GAGCTGCTCA GCAATACCA GCACAACGAG TCTCATCC CCCTGGCGAG
351 AGCATCCGCC AGGAGGAGCA AGGAGGAGAT CCTCATCTG CACAACAAGC
401 TTCGGGGCCA GGTGCAGCCT CAGGCCCTCCA ACATGGAGTA CATGACCTGG
451 GATGACGAAC TGGAGAAGTC TGCTGCAAGC TGGGCCAGTC AGTGATCTG
501 GAGCAGCGGC CCCAGCAGTC TGCTGTGCTC CATCGGGCAG AACCTGGCGC
551 CTCACTGGGG CAGGTATCGC TCTCCGGGGT TCCATGTGCA GTCTGTGAT
601 GACGAGGTGA AGGACTACAC CTACCCCTAC CCGAGCGAGT GCAACCCCTG
651 GTGTCCAGAG AGGTGCTCGG GGCCTATGTG CAGGCACCTAC ACACAGATAG
701 TTTGGGCCAC CACCAACAAG ATCGGTTGTG CTGTGAACAC CTGCCGAAAG
751 ATGACTGTCT GGGGAGAAGT TTGGGAGAAC GCGGTCTACT TTGTCTGCAA
801 TTAATCTCCA AAGGGGAAGT GGATTGGAGA AGCCCTCTAC AAGAATGGCC
851 GGCCTGCTCT TGAGTGCCCA CCCAGCTATG GAGGCACTG CAGGAACAAC
901 TTGTGTTACC GAGAAGAAAC CTACACTCCA AAACCTGAAA CGGACGAGAT
951 GAATGAGGTG GAACGGGCTC CCAATCTCTA AGAAACCAT GTTGGCTTCC
1001 AACGAGAGGT GATGACACCC ACCAAGGCCA AGAAACCTC TGGGTGAAC
1051 TACATGACCC AAGTCGTGAG ATGTGACACC AAGATGAAGG ACAGGTGCAA
1101 AGGGTCCACG TGTAAACAGT ACCAGTGCCC AGCAGGCTGC CTGAACCAAC
1151 AGGCGAAGAT CTTTGGAACT CTGTTCTATG AAAGCTCTGC TAGCATATGC
1201 GCGCGGCCCA TCCACTACGG GATCCTGAT GACAAGGAG CCGTGTGGA
1251 TATCAGCAGG AACGGGAGG TCCCTTCTT COTGAAGCT GAGAGACAGG
1301 GCGTCAGTC CCTCAGCAA TACAACCTT CCAGCTCATT CATGGTGTCA
1351 AAAGTGAAGG TGCAGGATT GGACTGCTAC ACGACGGTTG CTCAGTGTG
1401 CCGTTTGAA AAGCCAGCAA CTCACTGCCC AAGAATCCAT TGTCCGGCAC
1451 ACTGCAAGA CGAACCTTCC TACTGGGCTC CCGTGTGTTG AACCAACATC
1501 TATCAGATA CTTCAAGCAT CTGCAAGACA GCGGTGACG CCGGAGTCAT
1551 CAGCAACGAG AGTGGGGGTG ACGTGGAGCT GATGCCGTG GATAAAAGA
1601 AGACCTACGT GGGCTCGCTC AGGAATGGAG TTCAGTCTGA AAGCCTGGGG
1651 ACTCCTCGGG ATGGAAGGCC CTTCGGATC TTTGCTGTCA GGCAGTGAAT
1701 TTCCAGCACC AGGGGAGAAG GGGCGTCTTC AGGAGGCTT CCGGGTTTGG
1751 CTTTATTTT TATTTGTCA TTCCGGGTTA TATGGAGAGT CAGGAACCTT
1801 CTTTGACTG ATGTTCACTG TCCATCACTT TGTGGCTGT GGTGAGGTG
1851 ACATCTCATC CCCTCACTGA AGCAACAGCA TCCCAAGGTG CTCAGCCGGA
1901 CTCCCTGGTG CCTGATCTG CTGGGGCCCG GGGGTCTCCA TCTGACCTC
1951 CTCTCTCTCT TAGAGATCTG AGCTGTCTCT TAAAGGGGAC AGTTGCCCAA
2001 AATGTCTCTT GCTATGTGTT CTCTGTGG TGAGGAGAGT TGATTTCAAC
2051 CTCTCTGCTA AAAGAACAAA CCATTGGAAG CTCACAAATG TGAAGCATTC
2101 ACGGCGTCCG AACAGGCTTT TTGAGCAAG GCCAATGAGT TTCAGGAATG
2151 AAGTAGAAGG TAGTTATTTA AAAATAAAAA ACACAGTCCG TCCCTACCAA
2201 TAGAGGAAAA TGGTTTAAAT GTTTGCTGGT CAGACAGACA AATGGGCTAG
2251 AGTAAGAGGG CTGCGGGTAT CAGAGACCCC GGTTCGCCCC TGGCACGTGT
2301 CTTTCTGCTG GCGCGGCAC AGGCCCTT CAATGGCAGC ATTCAAGATG
2351 GCTCTATACA CAGCACTGCT GTTTATGTA GAGTTACGA GTCACTTCAG
2401 AGATGTATCT TGTCTTTGTC AGGCCCTTCA TCTTCATGCC CCACCTGTTT
2451 TCTGCCGTGA CTTTGTGTC CATTGAGGAC TAAGGATCGG GACCTTTCT
```

```
2501 TTACCCCTTA CCAATGTGG CTCCACCCCT GCCTGGGACT GGTTCACGTG
2551 TCCTGGTTCA CACCAGGAC TTTCTTTTC AAGCGAACCT GTTGAAGCC
2601 CAAGTCTTAA CTCCTGGTCT CGTAAGGTC CACTGAGACG AGATGTCTGA
2651 GAACAACCAA AGAAGGCGCT CTCCTTTGCT CTTTAAAAA ATGACAATTA
2701 AATGTGCAGA TTCCCGACGC ACCCGATGAC CTATTTTTTC AGCGTGGA
2751 GGAATGGAGT CTTTGTGACA TTCCCTACCG AGGTTAGCAG CTCAGTTTGT
2801 GGTATGAAA CCGTCTGTGG CCTCATGACA CGAGAGATG GGAATCACT
2851 AGAAGGATCT CTTTCTCTGT TTTCTGAAA CGACTCTTGC CAAACGTTCC
2901 CGAGGCGCCA AGGAGTGTAG TACACCTGG CTGCCATCAC TGTATAAAG
2951 TGCTTCATGA GCCCAGACCA AAGCCACACA GTGAATGAA GTACCTTTT
3001 GTAATAGCA TTTTGTGCA GAAGGTGAAA ATTCACCTCT CTACACCGG
3051 GCCAGCCAAT AGATCACTTT GGTGAATGCT AGTTTCAAT TTGATTCAAA
3101 ATATTCTTA GGTGAAGAA CTAGCAGAAA GTCAAAAAC TAAATACGT
3151 AGACTGGACA AGAATTCTA COTGGGCACC TAGTGATGC CTCTTTCTT
3201 TGATTGCTTT TCTAATAAT GCAGATCTG AAGTAAATA GCTTAAAC
3251 AAACAAAAA CCAACCCCTT TAAGAGTGTG GTAAAAAGCA GTTCACTCT
3301 TAGCTTGACT GAGCTAAAA TCAAGGACT ACCTGCTTG TGCATTGTAG
3351 TCTAGTCGA ATTATAGGT ACTGACTCT CAGCCCAAA TGTGGAGAG
3401 GAAGAATTCG GTCAGCCTGT CAGGTCTGTA GTCCAGTTAC CACCAACAT
3451 CTGGGAACCT TCTGGTCTGT GGTGCTCTG CTGCTGACT TTTGTGGCTG
3501 TGCTGTGTG TCAGAGATAA ATTAGATGC CTTGTGGGT TGCAGATT
3551 AGTGAAGGCT CCAGGACGAT CCAAGTGGG TCGCTTCAA AGCATCCAC
3601 TCAAGGAGA CTTGAACCT CAGGTGTGAG TTGACCCAT CATTTAAAA
3651 TAAAGTCCC GGTTCCTTA ATGCTCTCT CACTGGGCT TCTAGCAGG
3701 ATAGAAAGTC CTTGCCAGA GCAGGACCTG GCTGCTTTT TTTTITTTT
3751 TTTCCGAGA CCAAGTTCA CTCGTGTGC CAGGTAGAG TGCAGTGGC
3801 TGATCTCTGC TCATTGCAAC TGCCGCTCC CGGTTCAAG CAATTCTCAT
3851 GCATCAGCT CCCAATACC TGGGATACA GCGCTGAGT ACCATGCCCG
3901 GCTAATTTT GTATTTTAT TAGAGATGG GTTTCATTAT GTTGGCCAGG
3951 CTGGTCTCGA ACTCCTTACC TCAGGTGATC CACCCACCTT GGCCTCCGA
4001 AGTGTGGA TACAGGCTT CTTAGGAGT TTGCCCTACC GTATTCGAAA
4051 GTCCTTATCA TCCCAACAA CATTGTGAAA CTGGAATATT TGTCTTCAA
4101 AAATGGAAC AAGCATATA ATGATAAGC CTGCTCCAT CACCACTCT
4151 CCTGTGTG GAATAGAGG CCTCGTCT ACACACACT ACCCTGTGT
4201 TAAAAAGATC TTGTACCAAG CCAACGCGT TCTGGCTCT CTTGCCACA
4251 GGATGAACAT TTTCGGCTT CTTAGGAGT TTGCCCTACC GTATTCGAAA
4301 GCGTGTCTG GTTCTCTATA TTGCTGTAG GCTCACTAG CCGCAGTT
4351 ATGTGTCTG TTTTCTTAT GAAAAATCAT GTATTGTCT ACTTCTGTG
4401 TACAAAGTTT TATTGTAAAT GTTTTTGTG CTTGCATGA ACAGGGGCA
4451 CTTTGTGCA ATTGTTTCA TAGAAGTGT TTGATTCTA AATGTTCTT
4501 GTAACATAT TTTTATGAA AATCTGAAC AATTGTGAA ATAAACACT
4551 GAAACCAA AAAAAAAA AAAA
```

BLAST Results

Entry H5834352 from database EMBL:
human STS WI-15502.
Score = 1331, P = 5.4e-54, Identities = 287/301

Medline entries

98146272:
cDNA cloning of a novel trypsin inhibitor with similarity to
pathogenesis-related proteins, and its
frequent expression in human brain cancer cells.

Peptide information for frame 1

ORF from 205 bp to 1695 bp: peptide length: 497
Category: strong similarity to known protein

```
1 MSCVLGGVIP LGLLFLVCGS QGVLLPNVTL LEELLSKYQH NESHRSVRRA
51 IPREDKEEIL MLHNKLAGOV OPOASNMEYM TMDDELEKSA AAWASQIWE
101 HGPTSLIVSI GQNLGAHWGR YRSPGFHVQS WYDEVKDYTY PYPSECNPC
151 PERCSGPMCT HYTQIVWATT NKIGCAVNTC RKMTVMGEVM ENAVYFVCHY
201 SPKGNWIGEA PYKNGRPCSE CPPSYGGSCR NNLCYREETY TPKPETDEHM
251 EVETAPIEE NHWQLQPRVM RPTKPKTSA NWYTVQVRC DTKMKDRCKG
301 STCNRYQCPA GCLNKKAKIF GTLFYESSSS ICRAAIHYGI LODEKGLVDI
351 TRNGRVFFVF KSERHGVQSL SKYKPSSEFM VSKVKVQDL CYTTVAQLCP
401 FEKPATHCPR IHCPAHCKDE PSYMAPVFCT NIYADTSSIC KTAHVAGVIS
```

451 NESGGDVDM PVOKKTYVG SLRNGVQSES LGTPRDGKAF RIFAVRQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4b4, frame 1

TREMBL:AF109674.1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds., N = 1, Score = 968, P = 1.9e-97

TREMBL:D45027.1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds., N = 1, Score = 738, P = 4.5e-73

TREMBL:AB009609.1 gene: "HrTT-1"; Halocynthia roretzi HrTT-1 mRNA, complete cds., N = 1, Score = 345, P = 2e-31

PIR:JC5308 testis-specific, vespid, and pathogenesis-related protein 1 precursor - human, N = 1, Score = 337, P = 1.7e-30

>TREMBL:AF109674.1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds.

Length = 188

HSPs:

Score = 968 (145.2 bits), Expect = 1.9e-97, P = 1.9e-97
Identities = 160/185 (86%), Positives = 170/185 (91%)

Query: 61 MLHNKLRGOVOPQASNEHYMTWDDLEKSAWAASQCIWENGPTSLVLSIGQNLCAHMG 120
MLHNKLRGOV P ASNNEYMTWD+ELE+SAANA +C+WEHGP SLLVLSIGQNL HMGR
Sbjct: 1 MLHNKLRGOVYPPASNNEYMTWDEELERSAAWAQRCLEWENGPSLLVLSIGQLAVHMG 60

Query: 121 YRSPGFHVQSWYDEVKDYTYPPSECNWPCPERCSGPMCTHTYQIWMATTNKIGCAVNTC 180
YRSPGFHVQSWYDEVKDYTYPP ECNPWPCPERCSG MCTHTYQ+VMATTNKIGCAV+TC
Sbjct: 61 YRSPGFHVQSWYDEVKDYTYPPSECNWPCPERCSGAMCTHTYQHWATTNKIGCAVHTC 120

Query: 181 RKMTVWGEVWENAVYVFCNYSKGNWIGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY 240
R M+VMG+WENAVY VCNYSKGNWIGEAPYK+GRPCSECP SYGG CRNNLCYREE Y
Sbjct: 121 RSMVWGDWENAVYVFCNYSKGNWIGEAPYKHGRPCSECPSPSYGGSCRNNLCYREEHY 180

Query: 241 TPKEPE 245
KPE
Sbjct: 181 HQKPE 185

Pedant information for DKFZphtes3_4b4, frame 1

Report for DKFZphtes3_4b4.1

[LENGTH] 497
[MW] 55920.00
[pI] 8.36
[HOMOL] TREMBL:D45027.1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds. 6e-78
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YJL078c] 8e-12
[BLOCKS] BL01009E Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009D Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009C Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009A Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[PIRKM] glycoprotein 5e-22
[PIRKM] blocked amino end 5e-13
[PIRKM] brain 9e-30
[PIRKM] hydrolase 4e-09
[PIRKM] hemolymph coagulation 4e-09
[PIRKM] zymogen 4e-09
[PIRKM] alternative splicing 4e-09
[PIRKM] sperm 5e-22
[PIRKM] viroid-induced protein 2e-11
[PIRKM] venom 6e-18
[PIRKM] pyroglutamic acid 2e-11
[PIRKM] transmembrane protein 2e-10
[PIRKM] serine proteinase 4e-09
[SUPFAM] C-type lectin homology 4e-09
[SUPFAM] trypsin homology 4e-09

[SUPFAM] complement factor H repeat homology 4e-09
[SUPFAM] cysteine-rich secretory protein 1 6e-24
[SUPFAM] pathogenesis-related leaf protein 7e-15
[PROSITE] MYRISTYL 8
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 6
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 3
[PROSITE] SCP_AG5_PRI_SC7_2 1
[PFAM] SCP-like extracellular Proteins
[KW] All_Beta
[KW] SIGNAL PEPTIDE 23
[KW] LOW_COMPLEXITY 1.21 4

SEQ MSCVLGGVITPLGLFLVCGSQGYLLPNTLLLELLSKYQHNEHSRVRRAIPREDKEIL
SEGKXXXXX.....
PRD ccc

SEQ MLHNKLRGVQVQASINMEYMTWDOLEKSAAWASQCIMENGPTSLVSIQNLGAHWGR
SEG hhhhhhhcc
PRD ccc

SEQ YRSFGHVQSWYDEVKDTYTPYSECNPMCPERCSPGPKCTHYTIQVWATTNKIGCAVNTC
SEG
PRD ccc

SEQ RKMTVMGEVWENAVYFVCNYSFKGNWIGEPYKNGRPCSECPSPYGGSCRNNLCYRETY
SEG
PRD ccc

SEQ TPKPETDEHNEVETAPIPEENHVWLQPRVNRPTKPKKTSVNVYMTQVACDTKMKDRCKG
SEG
PRD ccc

SEQ STCNRYQCAGCLWKKAKIFGTLFYESSSSICRAAIHYGILDKGGGLVDITRNGKVPFFV
SEG
PRD ccc

SEQ KSERHGVQSLSKYKPSSTPMVSKVKVQDLDCYTTVAQLCFEKPATHCPRHCPAHCCKDE
SEG
PRD ccc

SEQ PSYWAPVFGTHIYADTSSICKTAVHAGVISNESGGVDVMPVDRKKTYYGSLRNGVQSES
SEG
PRD ccc

SEQ LGTPRDGKAFRIFAVRQ
SEG
PRD ccc

Prosite for DEF2phtes3_4b4.1

PS00001	27->31	ASN_GLYCOSYLATION	PD0C00001
PS00001	41->45	ASN_GLYCOSYLATION	PD0C00001
PS00001	451->455	ASN_GLYCOSYLATION	PD0C00001
PS00004	181->185	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	276->280	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	464->468	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	170->173	PKC_PHOSPHO_SITE	PD0C00005
PS00005	179->182	PKC_PHOSPHO_SITE	PD0C00005
PS00005	201->204	PKC_PHOSPHO_SITE	PD0C00005
PS00005	228->231	PKC_PHOSPHO_SITE	PD0C00005
PS00005	241->244	PKC_PHOSPHO_SITE	PD0C00005
PS00005	362->365	PKC_PHOSPHO_SITE	PD0C00005
PS00005	471->474	PKC_PHOSPHO_SITE	PD0C00005
PS00005	483->486	PKC_PHOSPHO_SITE	PD0C00005
PS00006	29->33	CK2_PHOSPHO_SITE	PD0C00006
PS00006	75->79	CK2_PHOSPHO_SITE	PD0C00006
PS00006	81->85	CK2_PHOSPHO_SITE	PD0C00006
PS00006	130->134	CK2_PHOSPHO_SITE	PD0C00006
PS00006	453->457	CK2_PHOSPHO_SITE	PD0C00006
PS00006	483->487	CK2_PHOSPHO_SITE	PD0C00006
PS00007	385->393	TYR_PHOSPHO_SITE	PD0C00007
PS00008	111->117	MYRISTYL	PD0C00008
PS00008	115->121	MYRISTYL	PD0C00008
PS00008	174->180	MYRISTYL	PD0C00008
PS00008	204->210	MYRISTYL	PD0C00008

WO 01/12659

PCT/IB00/01496

PS00008 227->233 MYRISTYL PDOC00008
 PS00008 300->306 MYRISTYL PDOC00008
 PS00008 447->453 MYRISTYL PDOC00008
 PS00008 470->476 MYRISTYL PDOC00008
 PS01010 195->207 SCP_AG5_PRI_SC7_2 PDOC00772

Pfam for DKFZphtes3_4b4.1

HMM_NAME SCP-like extracellular Proteins
 HMM *PQDEQDEWLNKHNDFRQOVGRGLETRGNPGPPASNMhPMVMNDELAT
 P + ++E+L HN +R QV P ASNM M+W+DEL +
 Query 52 PREDKKEILHLHNKLRGGVQ-----PQASHREYMTWDELEK 88
 HMM IAQNWANQCI FDHHDCCNHNhNPYQGNIAMWSsTANhPMhWssMIQNMWY
 A WA+QCI +H ++ + S GQN+ + + +++++ +Q+MY
 Query 89 SAAANASQCIhZHGPTSLVSI---GQNLGAHWG---RYASPGFHVQSWY 132
 HMM NEVLDYNYNHWTCkGG....NNFMVCGHYTONVMRnTfTIGCGRYICYC
 +EVKDY Y + + +C HYTO+VM+ T +IGC+ C+
 Query 133 DEVKDYTYPPSECNPMCPERCSPMCTHYTQIVWATTNKIGCAVNTCRK 182
 HMM HNNWcKDPWKhkhYYVCHVCPpGNtch*
 + M + M+ +Y VCNV P+GN+++
 Query 183 NTVM--GEVMENAVYFVCNYSKGNWIG 208

DKF1phtes3_4f17

group: testes derived

DKF1phtes3_4f17 encodes a novel 656 amino acid protein with weak similarity to methyl-CpG-binding proteins.

Methylation at the DNA sequence 5'-CpG is required for mammalian development. Methyl-CpG-binding proteins bind specifically to methylated DNA via a related amino acid motif and can repress transcription. The novel protein does not contain such a motif. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to methyl-CpG-binding protein

extension of HS557771/HS278337,
there are some differences to these sequences

Sequenced by AGOWA

Locus: /map="18"

Insert length: 2320 bp

Poly A stretch at pos. 2266, polyadenylation signal at pos. 2251

```
1 GGCAGGTTCC CGGTCGCTG CCGGGGGTCC TGAGGGAGTG CCGCGGGAGC
51 GGAGATATGG AGGGAGATGG TTCAGACCCA GAGCCTCCAG ATGCCGGGGA
101 GGACAGCAAG TCCGAGAATG GGGAGAATGC GCCCATCTAC TGCACTCGCC
151 GCAAACCGGA CATCAACTGC TTCATGATCG GGTGTGACAA CTGCAATGAG
201 TGGTTCCATG GGGACTGCAT CCGGATCACT GAGAAGATGG CCAAGGCCAT
251 CCGGGAGTGG TACTGTGGGG AGTGCAGAGA GAAGAGCCCC AACGTAGAGA
301 TTGGCTATCG GCACAAGAAG TCACGGGAGG GGGATGGCAA TGAGCGGGAC
351 AGCAGTAGGC CCGGGGATGA GGGTGGAGGG CGCAAGAGGC CTGTCCCTGA
401 TCCAGACCTG CAGCGCCGGG CAGGGTCAGG GACAGGGGTT GGGGCCATGC
451 TTGCTCGGGG CTCTGCTTCG CCCCACAAAT CCTCTCCGCA GCCCTTGGTG
501 GCCACACCCA GCCAGCATCA CCAGCAGCAG CAGAGAGAGA TCAAAAGGTC
551 AGCCCGCATG TGTGTGTGAG GTGAGGCATG TCGGCGCACT GAGGACTGTG
601 GTCACTGTGA TTTCTGTGCG GACATGAAGA AGTTCGGGGG CCCCACAAAG
651 ATCCCGCCAGA AGTGCCGGCT GCGCCAGTGC CAGCTCGGGG CCGGGGAATC
701 GTACAAGTAC TTCCCTTCCT CGCTCTCACC AGTACAGCCC TCAGAGTCCC
751 TGCCAAAGGC CCGCCGGCCA CTGGCCACCC AACAGAGGCC ACAGGCATCA
801 CAGAAGTTAG GCGCATCCG TGAAGTAGAG GGGGCACTGG CTTCTCAAC
851 AGTCAAGGAG CCTCCTGAGG CTACAGCCAC ACCTGAGCCA CTCTCAGATG
901 AGGACCTACC TCTGGATCCT GACCTGTATC AGGACTTCTG TGCAGGGGCC
951 TTTGATACCC ATGGCTTCGC CTGGATAGC GACACAGAG AGTCCCATTT
1001 CTTGGACCCC GCGCTCGGGA AGAGGGCAGT GAAAGTGAAG CATGTGAAGC
1051 GTCGGAGAA GAAGCTTGAG AAGAAGAAGG AGGAGCGATA CAAGCGGCAT
1101 CCGCAGAAAG AGAAGCACAA GGATAAATGG AATACCCAG AGAGGGCTGA
1151 TGCCAAGGAC CCTGCGTCAC TGCCCAAGTG CTTGGGGGCC GGCTGTGTGC
1201 GCCCGGCCCA CCCCAGCTCC AGTATATGCT CAGATGACTG TGCATGAAG
1251 TTGGCAGCCA ACCGCATCTA CGAGATCCTC CCCCAGGCGA TCCAGCAGTG
1301 GCAGCAGAGC CTTTGCATTG CTGAAGAGCA CCGCAGAAAG CTGCTGAAAC
1351 GCATTGCGCG AGAGCAGCAG AGTGCCCGCA CCGGCTTTCA GGAATGGAA
1401 CGCCGATTCC ATGAGCTTGA GGGCATCATT CTACGTGCCA AGCAGCAGGC
1451 TGTGGCGGAG GATGAGAGA GCAACAGAGG TGACATGAT GACACAGACC
1501 TGCAGATCTT CTGTGTTTCC TGTGGGCACC CCATCAACCC ACCTGTTCGC
1551 TTGGCGCACA TGGAGCGGTG CTACGCCAAG TATGAGGCC AGACGTCTCT
1601 TGGGTCCATG TACCCACAC GCATTGAAG GGGCACAGCA CTCTTCTGTG
1651 ATGTGTATAA TCCTCAGAGC AAAACATACT GTAAGCGGCT CCAGGTGCTG
1701 TGCCCGGAGC ACTCAGCGGA CCCCAGATG CACAGTGAG AGGTATGGG
1751 GTGCCCGCTT GTACCTGATG TCTTTGAGCT CACGGGTGAC TTCTGCCGCC
1801 TGCCCAAGCG CCAGTGCAT CCCCATTACT GCTGGAGAA GCTCGGGCT
1851 GCGGAAGTGG ACTTGGAGCG CTTGGCTGTG TGGTACAAGC TGGACGAGCT
1901 GTTTGAGCAG GAGCGCAATG TGCGCACAGC CATGACAAAC CCGCGGGGAT
1951 TCTTGGGCTT GATGCTGAC CAGAGATGC AGCAGATCC CCTCATACC
2001 GACCTGGGCT CAGTGCGGA CGGCTGAGCC TCTTGGCCCG GACCCCTTAC
2051 ACCCTGCAAT CCAGATGGG GAGCGCCCG GTGGCGGTGT GTCCGTCTCT
2101 CCACTCATCT GTTCTCCGG TTCTCCCTGT GGGCATCCAC CGGTTGACCG
2151 CCCATCTGCC TTTATCAGAG GGACTGTCCC CTTCCAGATG TTCAGTGCTT
2201 GGTGGGCTG CCGAGTCCAC TCATCTTCC CTCTCTCCC TGGGTTTGT
2251 TAATAAAATT TTGAAGAAC CAAAAAATA AAAAAAATA AAAAAAATA
2301 AAAAAAATA AAAAAAATA
```

BLAST Results

Entry H5557771 from database EMBL:ST:
Human chromosome 18 clone 2 mRNA sequence.
Score = 7582, P = 0.0e+00, identities = 1560/1598
Entry H5278337 from database EMBL:ST:
H.sapiens mRNA, expressed sequence tag ICRFp507H02194 (5')
Score = 6339, P = 9.0e-281, identities = 1307/1347
Entry H5095149 from database EMBL:ST:
human STS W1-6941.
Score = 1210, P = 2.2e-49, identities = 246/251

Medline entries

98449942:
Identification and characterization of a family of mammalian methyl-CpG
binding proteins.
9824997:
Gene silencing by methyl-CpG-binding proteins.

Peptide information for frame 3

ORF from 57 bp to 2024 bp: peptide length: 656
Category: similarity to known protein

1 MEGDGSQPEP PDAGEDSKSE NGENAPIYCI CRKPDINCFM IGCDCNENWF
51 MGDICIRITEK MAKAIREWYC RECREKDPKL EIRYRHKSR ERDGNERS
101 EPRDEGGGRK RVPVDPDLQR RAUSGTGVGA MLARGSASPH KSSPQLVAT
151 PSQHQQQQQ QIKRSARMCG ECEACRTED CGHCDPCRM KFGGFMKIR
201 QKRLAQCOL RARESYKYFP SLSPTVPSSE SLPRPRPLP TQOQPSOK
251 LGRIREDEGA VASSTVKEFP EATATPEPLS DEDLPDPL YDFCAGAFD
301 DHGLPMSOT EESFPLDPAL RKRAVKVHV KRREKSEK KEERYKRRHQ
351 KQKHKKWKH PERADAKDPA SLPCQLGPGC VRPAQPSKY CSDDCGMKLA
401 AMRIYEILPQ RIQWQOQSPC IAEHGHKLL ERIRREGQSA RTLOEHEAR
451 FHELEAITLR AKQAVAEDE ESHGOSDOT DLQITCVSGG RPIHPVALR
501 HNERCYAKYE SOTSFCSMPY TRIEGATRLF CDVYNPQSKT YCKRLQVLCF
551 EHSRDPKVPF DEVCGPLVR DVFELTGDFC RLPKRQCNH YCWEKLRRAE
601 VDLERVRVWY KLDELFEQER NVRTAMTNRA GLALMLHQT IQHDPLTTDL
651 RSSADR

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZp0034417, frame 3

TREMBL:CEFS2B11.4 gene: "F52B11.1"; Caenorhabditis elegans cosmid
F52B11, N = 2, Score = 316, P = 8.8e-27

TREMBL:HSAB2331.1 gene: "KIAA0333"; Human mRNA for KIAA0333 gene,
partial cds., N = 2, Score = 163, P = 2.8e-13

TREMBL:SPCC594.5 gene: "SPCC594.05c"; product: "putative
transcriptional regulatory protein, phd finger containing"; S.pombe
chromosome III cosmid c594., N = 3, Score = 168, P = 3.6e-12

TREMBL:AF072240.1 gene: "Mbd1"; product: "methyl-CpG binding protein
MBD1"; Mus musculus methyl-CpG binding protein MBD1 (Mbd1) mRNA,
complete cds., N = 2, Score = 189, P = 7.6e-11

>TREMBL:CEFS2B11.4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11
Length = 523

HSPs:

Score = 316 (47.4 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
Identities = 100/336 (29%), Positives = 167/336 (49%)

```
Query: 333 REKKSEKKKEERYKRRHQ-KQKHKKWHPERADAKDASLP-OCLGPGCVRPAPQSSKY 390
Sbjct: 118 QQRKANIINERDYPNAPTRQGSADLRARQTQNA-EPDKHPRQCLNPNCIYESRDSKY 176

Query: 391 CSDDCGMKLAANRIYEILPQRIQQW----QGSPTAEHGGKLLERIRREQSARTLQ 445
Sbjct: 177 CSDECGKELANRLTEILPNRCKQYFFEGPSGPRSLEDEIKPKRAKINREVQKLTSEK 236

Query: 446 EMERRFHEL-EAILRAKQAVREDEESNEGDSDDTLQIFCVSCGHPINPRVAL-RHME 503
Sbjct: 237 NMMAFLNKLVEFIKTQLKQLGTEERY-----DDNLYEGCIVGLPDIPLKRYTKHIE 290

Query: 504 RCYAKYESQTSFGSMYPTRIEGATRLFCQVYNPSQKTYKRLQVLCPEHSRDPKVPDEV 563
Sbjct: 291 LCMARSEKAIISFGA--PEK--NNDMFYCEKYSRTNSFCRLKSLCPEHRLGDEQHLKV 346

Query: 564 CGCP-----LVRDFELTQDF----CRLPKRCNRHYCEKLRRAEVDLVRV 607
Sbjct: 347 CGYPKKWEDQMISTAKTVSELIEMEDPFGEGRCKKDGACHKKHWIPSLRGITIELEQAC 406

Query: 608 VMYKLOELFEQ--ERNVRTAMTNRAGLALMLHOTIQHDFLTDLRSSA 654
Sbjct: 407 LFQKMYELCHEMHLNAHAEMTTNA--LSIMHMKQPSTEKCSFFLRNFA 453

Score = 53 (8.0 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
Identities = 24/100 (24%), Positives = 41/100 (41%)

Query: 169 CGCEACARRTEDCGHCDRCR----DMKK-FGGPNKIROKCRRLRQCOLRARESYKYPSS 222
Sbjct: 17 CMNCIRCNDKNCGTWPCRNKGTCDMRKCFSAKRLYNEKV-ROTDENLK-AIMAKTAQ 74

Query: 223 LSPVTPSESLPRPRPLPTQOOPQPSQKLRIR-EDEGAVASS 264
Sbjct: 75 REAAHQAAATTAPSAPVIEQVE-KKRGKRGSGNGGAA 116

Score = 48 (7.2 bits), Expect = 2.9e-26, Sum P(2) = 2.9e-26
Identities = 13/39 (33%), Positives = 19/39 (48%)

Query: 179 EDCGHCDRCRDMKKFGG--PNKIROKCRRLRQCOLRARESY 216
Sbjct: 15 ERCHNCIRCNDKNCGTWPCRNKGTCDMRKCFSAKRLY 53
```

Pedant information for DKFZphtes3_4f17, frame 3

Report for DKFZphtes3_4f17.3

```
[LENGTH] 656
[HW] 75711.71
[pI] 8.61
[HOWOLD] TREMBL:CEF52B11_4 gene: "F52B11.1": Caenorhabditis elegans cosmid F52B11 3e-25

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL138c] 3e-10
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL037c] 2e-04
[PROSITE] MYRISTYL 6
[PROSITE] AMIDATION 2
[PROSITE] CK2_PHOSPHO_SITE 8
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 9
[KW] All_Alpha
[KW] LOW_COMPLEXITY 18.75 %
[KW] COILED_COIL 4.57 %
```

```
SEQ MEGDGSDEPPDAGEDSKSENGENAPIYICIRKPDINCFMIGCDNCFNWFHDCIRITEK
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ NAKAIREMYCREKOPKLEIRYRHKKSREKRDGNERDSSEPRDEGGGRKRPVDPOLQR
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ RAGSGTGVGANLARGASAPHKSSPOPLVATPSQHQQOQQQIKRSARNCGEACARTED
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
```



```

SEQ      CGCHDCFRACGPGGPKK1RQCRLAQCOLAARESYKYVPSPLSVPTVSESLPRPRLPP
SEG      .XXXXXXXXXXXXXXXXXXXXX.XXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccc
COLLS    .

SEQ      TQOQPOPSQKLGRIREDQGAVSTVSEKPEATATPEPLSDPLDPLDYDFCAAGFD
SEG      .XXXXXXXXXXXXX.XXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COLLS    .

SEQ      DHGLPWSDTSESPLDALLRRAKVKVXVREKKSEAKKERYVRXROKQKDKRWKH
SEG      .XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COLLS    .

SEQ      PERADAKDPSLPQCLGPCVQAPQSPKSYSDQCGMKLANRRIYELPQRIQWQOSPC
SEG      .
PRD      hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhh
COLLS    .

SEQ      IAEHGKKLLERIREQOSARTLQCHERRFHELEAIRAKQAQVREDSHGSGSDS
SEG      .XXXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COLLS    .cccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      DLQIFCVSGCHPIPRVALRHMERCIYAKYESQTSFGSHYPTRIEGATLFCDVYVPSKT
SEG      .X
PRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccc
COLLS    .

SEQ      YCKRQLVQLCPHSRDPKVPADVEGGLPVRDVELTGDGFCRLPKRCRNYHCWEKLRAR
SEG      .
PRD      cccccchhhhhccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhh
COLLS    .

SEQ      VDLERRVRYWYKLDLFEQENRYVTAMTNAAGLMLMWLQIQTIDPLTDLRSSADR
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COLLS    .cccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_4f17.3

PS000002	124-1-128	GLYCOSAMINOGLYCAN	PGDC000002
PS000003	58-6-1	PKC_PHOSPHO_SITE	PGDC000003
PS000005	165-1-168	PKC_PHOSPHO_SITE	PGDC000005
PS000005	215-2-218	PKC_PHOSPHO_SITE	PGDC000005
PS000005	248-3-251	PKC_PHOSPHO_SITE	PGDC000005
PS000005	265-2-268	PKC_PHOSPHO_SITE	PGDC000005
PS000003	337-3-340	PKC_PHOSPHO_SITE	PGDC000003
PS000005	387-3-390	PKC_PHOSPHO_SITE	PGDC000005
PS000005	439-4-442	PKC_PHOSPHO_SITE	PGDC000005
PS000005	627-6-630	PKC_PHOSPHO_SITE	PGDC000005
PS000006	6-10-1	CK2	PGDC000006
PS000006	16-1-21	CK2	PGDC000006
PS000006	227-2-231	CK2_PHOSPHO_SITE	PGDC000006
PS000006	265-2-269	CK2_PHOSPHO_SITE	PGDC000006
PS000006	280-2-284	CK2_PHOSPHO_SITE	PGDC000006
PS000006	308-3-312	CK2_PHOSPHO_SITE	PGDC000006
PS000006	521-5-525	CK2_PHOSPHO_SITE	PGDC000006
PS000006	652-6-656	CK2_PHOSPHO_SITE	PGDC000006
PS000007	339-3-346	TYR_PHOSPHO_SITE	PGDC000007
PS000007	350-4-357	TYR_PHOSPHO_SITE	PGDC000007
PS000007	211-2-219	TYR_PHOSPHO_SITE	PGDC000007
PS000008	42-4-48	MYRISTYL	PGDC000008
PS000008	123-1-129	MYRISTYL	PGDC000008
PS000008	152-1-159	MYRISTYL	PGDC000008
PS000008	129-1-135	MYRISTYL	PGDC000008
PS000008	259-2-265	MYRISTYL	PGDC000008
PS000008	396-4-402	MYRISTYL	PGDC000008
PS000008	107-1-111	AMINOLIPID	PGDC000009
PS000009	425-2-429	AMINATION	PGDC000009

(No Pfam data available for DKFZphtes3_4f17.3)

DKFZphtes3_4f5

group: signal transduction

DKFZphtes3_4f5.3 encodes a novel 790 amino acid protein similar to beta-transducins.

The protein contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a cytochrome C family heme-binding site signature is present. The protein is larger (790 amino acids) than the usual eukaryotic G-beta transducins (about 340 amino acids).

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to S.pombe "beta-transducin"

complete cDNA, EST hits
complete cds,
on genomic level encoded by HS313D11, at least 7 exons these exons
match
only partially with the predicted transcripts in HS313D11

Sequenced by AGOWA

Locus: /map="16p13.3"

Insert length: 3166 bp

No poly A stretch found, no polyadenylation signal found

```
1 GCGGGCTTCC GCGCGGGCGG TTCCGGACAA CCGTGGCGTT TTAGTAAAG
51 ATTGGGGTTC GCGCGGGGGA GAAGGGCTGC CCGGGGCCCT CTGGTCTCTG
101 TCCCGCAGCG TCGCTCTCCC GCGGCCACTG CCGCGCTCCC AGGAACCTG
151 TACTCGGGGG TCGCGGGCTT CTCTCTGCGG TCGGGTCCGG CAGACACCT
201 CGAGCTCTTT AAGTAGCTCG GTCCCTTGACG TCCCTCTGGG CCGTCTCCCG
251 GTCTATCGCC TGAGTCCCGG GCGCCCTCTA GCGCTCTGTT CCGTCCCTTC
301 TTTTGTCTCT CCTAGAGCC CCGCGGCCCT CAGGGCTGAC AGTGTGGACG
351 GCGGGAGTCT CCGCTCTCCC CTGCTGGGAT TGACTGACCG AGCGTTTACT
401 GACTGGCCAG ATCTGGCTGA TGGGGGTACC GAGAGTGGC CTGGGCCCGG
451 AATGTCCAGC TAGAGCTTTC CGTGGAAATC AGACATGAAA CTGACAGGCC
501 TAAGGGAAGC TAGGAAGTCC CCGCACCCTG CAGCCAGGGT GATGGGCTGG
551 ACTGACAGAC TCCAGTGAAT TTGAGCTTGC CTGTCAAGCT GATTGGCTGA
601 TAGACAGCCC TCGATTGGCT CACTAAGACT GACCAAGCCG GGACCAAGCA
651 GTCTGGGGGT CCAACCTGGG GTGGAAAGTC TGAACCTATG ACCCACTCAG
701 GCTGACGAGG CCGACCCACC TCACTGACCT CCGTACCCCT GACCTCATCA
751 CCTGTGCAGC CATGGAGAAG ATGTCCCTGT TGACCACAGC CCGGGTGGGC
801 AGCGTGTCTG CAGGCGCGAC CATGCACTGC CACTGTGATG CTCCCGCCAA
851 TGCCATCAAT GTGTGCGCGG ACGACGCCCA GTGTGTCTGT GCGAGCGCTA
901 GCATCTTCAA GATCTATGCC ATCGAGGAGG AACAGTTGCT GGAAGAAGCTG
951 AACCTGCGTG TGGGGCGCAA GCCTTCGCTT AACCTGACCT GTGTGACGCT
1001 GGTCTGGCAC CAGATGGATG AGAACCTGCT GGCCACAGCA GCCACCAATG
1051 GCGTGGTGGT CAGGTGGAAC CTGGGCCGCG CATCCCCCAA CAGCAGGAC
1101 CAGCTGTCTA CAGACAGAAA GCGCACGGTA AACAAAGTCT GCTTCAACCC
1151 CACCGAAGCC CAGCTGCTGC TCAGTGGCTC CCAGGATGGC TTCATGAAGT
1201 GCTTTGACCT CCGCAGAAAG GACTCTGTCA GCACCTTCTC GGGCCAGTCTG
1251 GAGAGCGTGC GGGACGTGCA GTTCAGTATC CCGGACTACT TCACCTTCGC
1301 CTCCAGCTTT GAGAACGGCA ATGTGAGCT CTGGGACATC CGGCTGCGG
1351 ACCGCTGGCA GAGGATGTTT ACAGCCACCA ACGGACCGGT CTCTGCTGC
1401 GACTGGCACC CCGAGGACAG GGGCTGGTTG GCGACTGGAG GCGCGACAAA
1451 GATGGTGAAG GTCTGGGACA TGACCACGCA CCGTGGCAAG GAGATGCACT
1501 GTGTGCGAGC CATCGGCTCG GTGGCCGCTG TGAAGTGGGG CCGAGAGTGC
1551 CCGCACACCC TGCCACGCTG CTGCATGATG GTGGACACAA ACATCTATGT
1601 TTGGGACGTG CCGCGGCCCT TCGTGGCAGC TGCCATGTTT GAGGAACACC
1651 GAGAGCTCAC CACGGGAATT GCGTGGCGCC ACCCCACGCA CCGCTCTTTC
1701 CTGCTGTCTG GCTCCAAAGG CAGCTCGCTG TGCCAGCACC TGTTCGCGGA
1751 CCGCAGCCAG CCGCTCGAGC GCGCCAAACC TGAGGGCTTC TGTACGGCC
1801 TCTTGGGGGA CTGTGCTTTC GCGGCCAAGG AGAGCTTGTG GCGTGGCGAG
1851 TCGGGGCGCA AGCCCTACAC TGGCGACCGG CCGCCACCCA TCTTCTTTAA
1901 GCGCAAGCTG GACCTGCGCG AGCCCTTCCG AGGCTCTGCC TCGAGTGGCC
1951 TCAGTGTCTT TGAGACGGAG CAGGTGGGGG GCGGCATGCG CTGGTTTGTG
2001 GACACAGCTG AGCGTTATGC GCTGCTGGCG GCGGCATGCG CCGAGCTCTG
2051 TGACACACAC GCAAGAGTGG CTGAGAGACT TGCGCGCAAC CAGGTGGCGG
2101 AAACGTGGAC CATGCTGGCG ATCATCTACT GCAGCCCTGG CCTAGTGGCC
2151 ACTGCAAAAC TCAACACAGG TGTGGGCAAG GTGGGCTCCT GTGGCTCTCC
2201 GCTCATGAAC AGTTTCAACC TGAAGGATAT GCGCCACAGG TTGGGCAATG
2251 AGACGGGCTT GGACCGAGC AAAGGAGATG CAGCGAGGCA CACAGTTCTG
2301 CTGCACTCTT CCGCCACACT CATACCAAT GAGGATAACG AGGAACCGGA
2351 GGGCAGCGAC GTACCTGCGG ACTACCTGAT GGGTACAGTG GAAGGTGAGG
```

```
2401 AGGACGAGCT GTACCTGCTG GATCCGGAAAC ACCGGCACCC CGAGGACCOCT
2451 GAGTCCCTCC TGCCCGCAGGA GGCTTTTCGG CTGGCCACAC AGATCGTGGA
2501 CACGCCCTCCC GGACCCGAGC ACCTGCAGGA CAAGGCCGAC TCCTCCGCAGC
2551 TGAGCCGCAG CGAGCGGAT GTGGCTCCG TGGCCCGCT GGACTCTCC
2601 TTCTCGCTCC TGTCTGCTC ACACGGCTC TACGACGCC GCCTGCCGCC
2651 CGACTTCTTC GCGCTGCTGG TGGCGGCAT GCTGCACCTC TACGCTGAGC
2701 AGGGCGACCT GCAGATGGCT GTGTCTGTC TCATCTCTCT GGGTGAACGG
2751 GTGGCGAAGG ACATCGACGA GCAGACCCAG GAGCACTGCT ACATCTCTTA
2801 CATGCACCTG CTGCAGCCT TCCGCTCTG GAGCTGTCC AAGGAGTGG
2851 TCAAGCTGAG CACCAGCCGC GCGCTCAGCT GCCTCAACCA GGCCTCCACC
2901 ACCCTGCAGG TCAACTGCAG CCACTGCAG GCGCCATGA GCAGCCGGGG
2951 CTGGGTCTGC GACAGTGCC ACCGCTGCG CAGCATGTGT GCGCTCTGCC
3001 ACCACGTAGT CAAGGGTCTC TTCTGTGCT GCAAGGGCTG CAGCCAGCG
3051 GGCACCTGCG ASCACATCAT GAGCTGGCT GAGCGAGCT CCACTCTCC
3101 CGAGGCTGCG GCGCACCTCT GCGAGTACTC CTGACGGGGC ATCTGCTGGG
3151 CTTGCCCGGG CGGCCG
```

BLAST Results

Entry HS313D11 from database EMBL:
Human DNA sequence from cosmid 313D11 from a contig on the short arm of
chromosome 16. Contains ESTs, STS and CpG islands.
Score = 6238, P = 0.0e+00, identities = 1318/1391

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 762 bp to 3131 bp; peptide length: 790
Category: similarity to known protein

```
1 MEKMSRVTTA LGGSVLTGRT MHCHLDAPAN AISVCRDAAQ VVVAGRSIFK
51 IYAIIEEQFV EKLNLVGRK PSNLNSCADV VWHOMENLL ATAATNGVVV
101 TWNLRPSRM KQDLITENK RYVKEVCFHP TEARVLSGS ODGPNKEFDL
151 RKQSVSTFS QGSESVRQV PSIRDYITFA STFENGIVQL WDIRRPDCE
201 RMFTAHNGPV FCCDWHFEDR GMLATGGDK MVKWMQMT RAKENHCVQT
251 IASVARVWR PECDRHLLATC SHMVHNIIV MDVRRPFVPA AMFEHRDVT
301 TGIAMRHFD PSFLLSGSKD SSLCOHLFRD ASQVPERAN EGLCYGLFUD
351 LAFAAKESLV AESGRKPYT GORNRPIFK RLIDPAEPA GLASSALSUF
401 DTEPGGGGHR WFYDTAERYA LAGRPLAELC DHNAKAREL GRNQAGTWT
451 MLRIIYCSPG LVPTANLHNS VCKGSGCGLP LMSFNLMOM APGLGSETRL
501 DRSGDARS D TVLLDSSATL ITNEDMEETE GSDVPADYLL GDVEGEDEEL
551 YLLOPEHARP EDPECVLPQE AFLRHEIVD TPPEPEHLD KADSPHVSQS
601 EADVASLAPV DSSVLSLVS HALYDRLPP DFTQVLVRDM LHFYASQGV
651 QMAVSVLIVL GERVRKIDE QTOEHMYSY IDLQRFRLM NVSNEVVKLS
701 TSRAVSLMQ ASTTLHVNCS KCRPMSSRG WVCORCHCA SMCVACHHV
751 KGLFVVCQG SHGGHLQHIM KWLEGSSHCP AGCCHLCEYS
```

BLASTP hits

Entry YDSB_SCHPO from database SWISSPROT:
HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN
CHROMOSOME 1. >TREMBL:SPAC4F8.11 gene: "SPAC4F8.11"; product:
"beta-transducin"; S.pombe chromosome 1 cosmid c4f8.
Score = 404, P = 3.0e-42, identities = 169/639, positives = 278/639

Entry PEX7_HUMAN from database SWISSPROT:
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
>TREMBL:HSU76560.1 gene: "Pex7"; product: "peroxisome targeting signal
2 receptor"; Human peroxisome targeting signal 2 receptor (Pex7) mRNA,
complete cds. >TREMBL:HSU88871.1 gene: "HsPEX7"; product: "HsPex7p";
Human HsPex7p (HsPEX7) mRNA, complete cds.
Score = 220, P = 1.1e-15, identities = 62/244, positives = 107/244

Entry PEX7_MOUSE from database SWISSPROT:
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
>TREMBL:M9171.1 product: "peroxisomal PTS2 receptor"; Mus musculus
peroxisomal PTS2 receptor mRNA, complete cds.
Score = 214, P = 5.3e-15, identities = 60/240, positives = 106/240

WO 01/12659

PCT/IB00/01496

Entry ATAC2294.7 from database TREMBL:
 gene: "FliP17.7", Arabidopsis thaliana chromosome I BAC FliP17 genomic
 sequence, complete sequence.
 Score = 232, P = 3.4e-14, identities = 68/260, positives = 120/260

Entry S66835 from database PIR:
 probable membrane protein YOL138c - yeast (Saccharomyces cerevisiae)
 >TREMBL:SCYOL138C_1 S.cerevisiae chromosome XV reading frame ORF
 YOL138c
 Score = 136, P = 2.5e-13, identities = 24/77, positives = 44/77

Alert BLASTP hits for DKFZphtes3_4f5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_4f5, frame 3

Report for DKFZphtes3_4f5.3

```

[LENGTH]      790
[RM]           88207.10
[PI]           6.05
[MOMOL]        SWISSPROT:YDSB_SCHPO HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN
C4FB.11 IN CHROMOSOME I. 9e-44
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YOL138c] 5e-16
[FUNCAT]       10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 3e-11
[FUNCAT]       3e-10
[FUNCAT]       04.05.01.01 general transcription activities [S. cerevisiae, YBR198c]
TAF90 - TFIID subunit] 9e-09
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YLL011w] 1e-07
[FUNCAT]       3e-07
[FUNCAT]       30.09 organization of intracellular transport vesicles [S. cerevisiae,
YDL195w] 2e-07
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL195w]
2e-07
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YDR142c] 4e-07
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]
4e-07
[FUNCAT]       08.10 peroxisomal transport [S. cerevisiae, YDR142c] 4e-07
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       04.07 rna transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YER107c] 4e-07
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT]       06.13 proteolysis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 8e-07
[FUNCAT]       04.05.03 rna processing (splicing) [S. cerevisiae, YPR178w] 1e-06
[FUNCAT]       03.13 meiosis [S. cerevisiae, YLR129w] 3e-06
[FUNCAT]       03.25 cytokinesis [S. cerevisiae, YCR057c] 1e-05
[FUNCAT]       1e-05
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 2e-04
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YOR272w] 6e-04
[SCOP]         dlgothb_2.46.3.1.1 betal-subunit of the signal-transducing 5e-06
[PIRKW]        duplication 7e-10
[PIRKW]        signal transduction 7e-08
[PIRKW]        peroxisome 9e-06
[PIRKW]        heterotrimer 7e-08
[PIRKW]        GTP binding 7e-08
[PIRKW]        peroxisome biogenesis 9e-06
[PIRKW]        transmembrane protein 1e-14
[SUPFAM]       MS11 protein 7e-10
[SUPFAM]       WD repeat homology 1e-14
[SUPFAM]       GTP-binding regulatory protein beta chain 7e-08
[SUPFAM]       PR11 protein 3e-08
[SUPFAM]       coatamer complex beta' chain 1e-06
[PROSITE]      CYTOCHROME_C 1
[PROSITE]      WD_REPEATS 3
[PROSITE]      MYRISTYL 10
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 11
  
```

```

(PROSITE) TYR_PHOSPHO_SITE 1
(PROSITE) PKC_PHOSPHO_SITE 7
(PROSITE) ASN_GLYCOSYLATION 4
(PFAM) WD_domain, G-beta repeats
(KW) Ali_Beta
(KW) 10
(KW) LOW_COMPLEXITY 2.28 %

```

```

SEQ MEKMSRVTTALGGSVLTGRTHCHLQAPANAISVCRDAAQVVVAGRSIFKIYAIIEEFQV
SEG .....
IgotB .....
SEQ EKLNLRVGRKPSNLNSCADVVMQMDENLLATAATNGVVVTWNLGRPSRNKQDQLFTEHK
SEG .....
IgotB .....TTCEEEEEETTTTTEEEET-TTCEEE--EEECCE
SEQ RTVNKVCFHPTFAHVLGSGQDGFNCFDLARKDSVSTFSGQSESVRDVQFSIRDYITFA
SEG .....
IgotB .....CCEEEEEETTTTCEEEETTTTTEEEETTTTTEEEECBVTCEEEEEETTTTTEEE
SEQ STFENGVLQMDIRRPORCERMTAHRNGPVFCCDWHPEDRGLATGGRDKMKVVMWMTTH
SEG .....
IgotB .....E-ETTTTEEEETTTTTEEE-EEECCEEEEEEC-TTTTCEEEEEETTTTTEEE
SEQ RAKENHCVQTIASVARVWRPECRHHLATCSMHVDHNIYVMDVRRFPVPAAMFEHRDVT
SEG .....
IgotB .....
SEQ TGIAMRHDPHPSFLLSGSKDSSLCQHLFRDASQPERANPEGLCYGLFGDLAFAAKESLV
SEG .....
IgotB .....
SEQ AAESGRKPYTGDRHPIFFKRLDPAEPFAGLASSALSVFETEPGGGHRMFVDTAERYA
SEG .....
IgotB .....
SEQ LAGRLAELCOHNAKVAELGRHQVAQTWNLRIIYCSPLVPTANLHNSVKGGSCLLP
SEG .....
IgotB .....
SEQ LKNSFNLRDMAPGLGSETRLDKSGDARSQVLLQSSATLITNEONEETEGSDVPADYLL
SEG .....
IgotB .....
SEQ GDVEGCEDELYLLDPEHARPEDECVLQEAFFLAHEIVDTPGPEHLQDKADSPHVSQS
SEG .....
IgotB .....
SEQ EADVASLAPVDSSFSLLSVSHALYDSRLPDDFFGLVRDMLHFAEGGDVQMAVSVLIVL
SEG .....
IgotB .....
SEQ GERVRKIDIEQTEHMYTSYIDLLQRFRLMNSNEVVKLSTSRVSCNLQASTTLHVNCS
SEG .....
IgotB .....
SEQ HCKRPMSSRGWVCDRCRCASMCVCHVVGLEFVWCQGCSSHGGLQHIMXWLGSSHCP
SEG .....
IgotB .....
SEQ AGCGHLCEYS
SEG .....
IgotB .....

```

Prosites for DKFZphtes3_4f5.3

PS00001	74->78	ASN_GLYCOSYLATION	PD0000001
PS00001	468->472	ASN_GLYCOSYLATION	PD0000001
PS00001	691->695	ASN_GLYCOSYLATION	PD0000001
PS00001	718->722	ASN_GLYCOSYLATION	PD0000001
PS00004	69->73	CAMP_PHOSPHO_SITE	PD0000004
PS00004	152->156	CAMP_PHOSPHO_SITE	PD0000004
PS00005	17->20	PKC_PHOSPHO_SITE	PD0000005
PS00005	165->168	PKC_PHOSPHO_SITE	PD0000005
PS00005	172->175	PKC_PHOSPHO_SITE	PD0000005
PS00005	239->242	PKC_PHOSPHO_SITE	PD0000005
PS00005	364->367	PKC_PHOSPHO_SITE	PD0000005
PS00005	701->704	PKC_PHOSPHO_SITE	PD0000005

WO 01/12659

PCT/IB00/01496

PS000005	727->730	PKC_PHOSPHO_SITE	PD0C000005
PS000006	76->80	CK2_PHOSPHO_SITE	PD0C000006
PS000006	165->169	CK2_PHOSPHO_SITE	PD0C000006
PS000006	172->176	CK2_PHOSPHO_SITE	PD0C000006
PS000006	181->185	CK2_PHOSPHO_SITE	PD0C000006
PS000006	398->402	CK2_PHOSPHO_SITE	PD0C000006
PS000006	498->502	CK2_PHOSPHO_SITE	PD0C000006
PS000006	503->507	CK2_PHOSPHO_SITE	PD0C000006
PS000006	522->526	CK2_PHOSPHO_SITE	PD0C000006
PS000006	598->602	CK2_PHOSPHO_SITE	PD0C000006
PS000006	600->604	CK2_PHOSPHO_SITE	PD0C000006
PS000006	675->683	CK2_PHOSPHO_SITE	PD0C000006
PS000007	337->346	TYR_PHOSPHO_SITE	PD0C000007
PS000008	13->19	MYRISTYL	PD0C000008
PS000008	97->103	MYRISTYL	PD0C000008
PS000008	139->145	MYRISTYL	PD0C000008
PS000008	161->167	MYRISTYL	PD0C000008
PS000008	317->323	MYRISTYL	PD0C000008
PS000008	342->348	MYRISTYL	PD0C000008
PS000008	391->397	MYRISTYL	PD0C000008
PS000008	460->466	MYRISTYL	PD0C000008
PS000008	474->480	MYRISTYL	PD0C000008
PS000008	759->765	MYRISTYL	PD0C000008
PS000009	67->71	AMIDATION	PD0C000009
PS000009	364->368	AMIDATION	PD0C000009
PS00190	743->749	CYTOCHROME_C	PD0C00169
PS00678	90->105	WD_REPEATS	PD0C00574
PS00678	223->238	WD_REPEATS	PD0C00574
PS00678	269->284	WD_REPEATS	PD0C00574

Pfam for DKFZphtes3_4f5.3

HHH_NAME	WD domain, G-beta repeats		
HHH	*MKGHndMVMCVaFSPDGtWFiVSGSMDgTCRLWD*		
	** HN+V C+ ++P+ R +++G+D+ +++WD		
Query	203	FTAHNGPVFCDDMHPEDRGWLATGGRDKMKVWVD	236

DKFZphtes3.4h6

group: intracellular transport/trafficking

DKFZphtes3.4h6 encodes a novel 622 amino acid protein with strong similarity to the kinesin light chain.

kinesin is a microtubule-based motor protein that pulls vesicles or organelles towards the plus end of microtubules. Structural changes in the protein that drive motility are coupled to ATP binding and hydrolysis. The novel protein is similar to kinesin light chain, which is part of the functional kinesin holoenzyme tetrameric protein. The light chain has been proposed to function in coupling of cargo to the heavy chain or in the modulation of the ATPase activity of the heavy chain. The novel protein contains two kinesin light chain repeats and one RGD cell-attachment site.

The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

strong similarity to Kinesin light chain

complete cDNA, complete cds, start at 150, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 2992 bp

Poly A stretch at pos. 2914, polyadenylation signal at pos. 2893

```
1 GCGCGGATGG AGCGGGCGGG ACCGGCTCGC GGGTGGGGGT CCGGGTGAAG
51 CGGGAGGCGAG CCAGAGTCGG AGCGGGGCCC GAGCACCAGG CGCAGGCCCG
101 GCGCCCGCCT GCCCGCACCC TCGTCCTCAC AGACGCCACA GCCATGGCCA
151 TGATGGTGTG TCCCGGGGAG GAGAAGCTGA GCCAGATGA GATCGTGTG
201 GGCACCAAGG CTGTATCA CAAGACTGGG ACTCTGCTG GGAACATG
251 TGCCCTGTGT GCTCTCTGCT TTGACCTGTA GCGCGGCGAA CGCGAGCCTG
301 GCTCGCAGGA GCGCTGCATC CTCTGCGTCT GCTCCCTGGA AGCCATTGAG
351 CTGGGGCTGG GGGAGGCCCA GGTGATCTTG GCATTGTGGA GCCACCTGGG
401 GGCTGTAGAA TCAGAGAAAC AGAAGCTGGC GCGCAGGTG CCGCTGTCTG
451 TCGAGGAGAA CCACTGCTGT CTGAGGAGC TGGCGGGGAC ACAGCAGAG
501 CTGACGCGCA GTAGCAGGC CTTGGCCGAC CTGAGGAGG AGAAGCAGCA
551 CTGTCTGTTT ATGAGCCAGA TCCGCAAGTT GGATGAAGAC GCTCCCTCTA
601 ACGAGGAGAA GGGGACGCTC CCCAAGACA CACTGGATGA CTTGTTCCCC
651 AATGAGGATG AGCAGAGCCC AGCCCTTACG CCAGAGGAG GGAATGTGTC
701 TGCTCAGCAT GGGGGCTAGC AGATCCCGCG CCGGCTCCCG ACCCTGCACA
751 ACCTGGTGAT CCAATACGCC TCACAGGGCC GCTACGAGT AGCTGTGCCA
801 CTCTGCAAGG AGGCACTCGA AGACCTGGAG AAGACGTGAG GCCACGACCA
851 CCCTGACGTT GCCACCATGC TGAACATCTT GGCACGTGTC TATCGGGATC
901 AGAACAGTGA CAAGGAGGCT GCGCACCTGC TCAATGATGC TCTGGCATC
951 CGGGAGAAJA CACTGGGCAA GGACACCCCA GCGGTGGCTG CGACACTAAA
1001 CAACCTGGCA GTCTGTATG GCAAGAGGGG CAAGTACAAG GAGGCTGAGC
1051 CATTGTGCAA GCGGGCACTG GAGATCCGGG AGAAGGTCTT GGGCAAGTTT
1101 ACCCAGATG TGGCCAGCA GCTCAGCAAC CTGGCCCTGC TTGTCAGAA
1151 CACGGGCAAA GGTGAGAGG TGGATATTA CTATCGCGCG CCACTGAGGA
1201 TCTATGCTAC ACGCTCGGG CCGATGACC CCAATGTGGC CAAGACCAAG
1251 AACCACTGGG CTCTCTGCTA CCTGAAGCAG GCAAGTACC AGGATGCGGA
1301 GACCTTGTAC AAGGAGATCC TCACCGCGCG TCATGAGAAA GAGTTTGGCT
1351 CTGTCAATGG GAGCAACAG CCCATCTGGA TGACGCGA GAGCGGGAG
1401 GAAAGCAAGG ATAGCGCCCG GGACAGCCCG CCCTATGGGG AATACGGCAG
1451 CTGGTACAAG GCCTGTAAAG TAGACAGCCC CACAGTCAAC ACCACCTGCG
1501 GCAGCTTGGG GGCCTATATC CGCGCCGAGG GCAAGCTGGA AGCCGCGCAC
1551 ACACTAGAGG ACTGTGCCAG CCGTAACCGC AAGCAGGTTT TGGACCCCGC
1601 AAGCCAGACC AAGTGTGAG AACTCTGAAA AGATGGCACT GGCAGGGGGG
1651 GAGACCGCCG CAGCAGCCGA GACATGGCTG GGGGTGCCGG GCCTCGGTCT
1701 GAGTCTGACC TCGAGGAGCT GGAACCTACA CTTGAGTGA ATGGGGATGG
1751 CAGTGGCTCC TTGAGGCGCA CGGTTCTCTT TGGGAACTC CGGATGCCCC
1801 TGAGGCGCAG CAGTGAGATG CTGTAAAGA AGCTGAGGG GGGACCCCCC
1851 CAGGAGCCCC CTAAACCCAG GATGAAGCGG GCACTGTCCG TCACTCTCTC
1901 CAACAAGAGC GTGAAGAGC GCACCCAGCC TGAAGGCACA GTTCTCTCTG
1951 ACAGCCGAC TCTCAGCTCC AGCTCCATGG ACCTCTCCCG ACGAAGCTCC
2001 CTGGTGGGCT AATGCTGAAG GGGCAGCCAG TCACGAGAGC GCCACCTGG
2051 CACACCCCCC TCACCCGAGC CTGCGCATG GAGCTGTCTG TTGTCCGCGC
2101 TGTCTCTCCG ACAGCCCTG TCTTTCTCTG TCAATCTCAG GGTAACTTTC
2151 TCCCTTGTCA TCTCAGCTG AGCCCTGGAG GCTGGCCCTG CCCACTCCAG
2201 CTCCATCCCT TATTATTCCC TTCCAGAGG GCGCTTCTCC CTAGGTTCCG
2251 GCCAGCAGGA GGTGCGGCTT GAGTGTCCA CCAATAGACT AGTGGCCTGG
2301 CTTCCGCGCA CCCCAGAGC AAGAACCTA AGCACTCCG GCGCTCTGG
2351 CACCTTCGCC CTCCCTCCG ACTCAACCG GCGGTGCTT CTGTATATAG
2401 AGAAATAAGT TATTGCCCCG GCGCTCCCT TCACTCCAG GTACTACCG
```

WO 01/12659

PCT/IB00/01496

```
2451 GGCTCCGCTT CGTCCCTCTT CTAGTGGTAC CGCCAGGCC TTAATCACCC
2501 CCATTCCGTT CGTCCGATATC TCCAGGCTC TACATTCTCG GGAGCGCGCC
2551 CTCCCAAGGG GGTCCCTGGG CCTTCTCGG CTCTCTCTGG CCTCTGAGGG
2601 ATGCGTCTTA CCGCGCCAT CCGCCCTGG CCGAGAGGG GGAGCTCCGC
2651 TTAGTCCGTC CTCCACCGCC CGGCGCTTC CCGCATCCG GGCCTTATGC
2701 ACTGCCCTTC CCACCCGGCC CGGCCAGGC AGGCCGACC CGGCCCGGGG
2751 CACCGCCAC CGAGCCATCC TGGCTCGCT CCGCCAGCC CTGCAGCTTC
2801 TCGCGAGGGG CGGCGACGGT CCGCTGGTG CAGGAGGGG TCCCTCTGTT
2851 CGGGGTGAGG CGCTGCTCT CTATTTTCA AGTTGCTGT AGAATPAAAG
2901 ACGGTTTAAA TCTGAAGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA
2951 AAAAAAAAA AAAAAAAAA AAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

9828268:
Two kinesin light chain genes in mice. Identification and
characterization of the encoded proteins.

Peptide information for frame 3

ORF from 144 bp to 2009 bp; peptide length: 622
Category: strong similarity to known protein
Prosite motifs: RGD (502-505)
KINESIN_LIGHT (223-265)
KINESIN_LIGHT (265-307)

```
1 MANMVFPRER KLSQDEIVLG TKAVIQGLT LRGEHRALLA PLVAPEAGEA
51 EPGSQERCIL LRSLEAIEL GLGEAQVILA LSSHGAVES EKQKLAQVR
101 RLVOENQWLR EELAGTQOKL QRSEQAVAGL EEEKQHLIFM SQIRKLDEDA
151 SPNEEKGDVP KDTLDDLPFN EDEQSPAPSP GGGDVSQGHG GVEIPARLRT
201 LHMVVIQYAS QGRYEVAVPL CKQALEDEX TSGMDHPVA TMLNIALVY
251 RQHWYKEAA HLLNDLAIH ETLCKDMPA VAANTLNIAV LYKRGQRYKE
301 AEPFLCKRALE IREKVLGKFH PDAKQLSML ALLCQNGKA EEEYYYRRA
351 LEIYATRLGP DDPNVARTKN NLASCYLKQG KYQDAETLYK EILTRAHEKE
401 FGSVNGDNKP IWMHAEREE SKDKRRDSAP YGEYGSWYKA CKVDSPTVNT
451 TLASLGALYR ROKLEAANT LDCASRRRK QGLPASQTK VVELLDGSG
501 RRGDRRSSRD NAGGACPRSE SDLEDVCPA EWNGDGSGL RRSQSFGLR
551 DALRRSSEML VKKLQGGTPO EPPNPRMKRA SSLNFKNSV EEPQPGGTG
601 LSDSRTLSSS SMDLSRRSSL VG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZp383_4h6, frame 3

TREMBL:AF055666.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus
musculus kinesin light chain 2 (Klc2) mRNA, complete cds., N = 1, Score
= 2824, P = 4e-294

PIR:I53013 kinesin light chain - human, N = 1, Score = 1927, P =
4.5e-199

PIR:C41339 kinesin light chain C - rat, N = 1, Score = 1919, P =
3.2e-198

SWISSPROT:KNLC_RAT KINESIN LIGHT CHAIN (KLC), N = 1, Score = 1919, P =
3.2e-198

>TREMBL:AF055666.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus
musculus kinesin light chain 2 (Klc2) mRNA, complete cds.
Length = 599

HSPs:

Score = 2824 (423.7 bits), Expect = 4.0e-294, P = 4.0e-294
 Identities = 558/598 (93%), Positives = 572/598 (95%)

```

Query:   1 NAMNVFPREKLSQDEIVLGTFAVIOGLETLAGENRALLAPLVAPEAGEAEFGSQERCIL 60
NA NV PREKLSQDEIVLGTFAVIOGLETLAGENRALLAPLVAPEAGEAEFGSQERCIL
Sbjct:   1 NATHVLPREKLSQDEIVLGTFAVIOGLETLAGENRALLAPLASHEAGEAEFGSQERCIL 60

Query:   61 LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL 120
LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL
Sbjct:   61 LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL 120

Query:   121 QRSEQVAQLEEEKQHLFMSQIRKLDEADSPNEEKGDVPKDTLDDLPNEDEQSPAPSP 180
QRSEQVAQLEEEKQHLFMSQIRKLDE PEEKGDVPKD+LDDLPNEDEQSPAPSP
Sbjct:   121 QRSEQVAQLEEEKQHLFMSQIRKLDE-MLPQEEKGDVPKDSLDDLPNEDEQSPAPSP 179

Query:   181 GGGDVSGQHGGEYI PARLRTLNHLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 240
GGGDV+ QHGGEYI PARLRTLNHLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA
Sbjct:   180 GGGDVAAQHGGEYI PARLRTLNHLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 239

Query:   241 TMLNIALVYRDQNKYKAAHLNDALAIREKTLGKDHFAVAATLNHLAVLYGKRGKYKE 300
TMLNIALVYRDQNKYK+AAHLNDALAIREKTLGKDHFAVAATLNHLAVLYGKRGKYKE
Sbjct:   240 TMLNIALVYRDQNKYKAAHLNDALAIREKTLGKDHFAVAATLNHLAVLYGKRGKYKE 299

Query:   301 AEPCLCKRALEIREKVLGKFPDPAKQSLNALLCONQKAEVEYYRRALEIYATRLGP 360
AEPCLCKRALEIREKVLGKFPDPAKQSLNALLCONQKAEVEYYRRALEIYATRLGP
Sbjct:   300 AEPCLCKRALEIREKVLGKFPDPAKQSLNALLCONQKAEVEYYRRALEIYATRLGP 359

Query:   361 DDPNVAKTNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNGKPIWMHAEREE 420
DDPNVAKTNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNG+KKPIWMHAEREE
Sbjct:   360 DDPNVAKTNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNGKPIWMHAEREE 419

Query:   421 SKDKRRDSAPYGEVGSYKACKVDSPTVNTTLASLGALYRQCKLEAANTLEDCASRRK 480
SKDKRRD P EYGSYKACKVDSPTVNTTLR+LGALYR +GLEAANTLEDCASR+RK
Sbjct:   420 SKDKRRDRPM-EYGSYKACKVDSPTVNTTLRLTGLALYRPEGKLEAANTLEDCASRRK 478

Query:   481 QGLDPASQTKVVELLKDGSGRRSSRDMAAGGAPRSESDLEOVGPTAEWNGDGSGL 540
QGLDPASQTKVVELLKDGSGR G RR SRD+AG P+SESDLE+ GP AEW+GDGSGSL
Sbjct:   479 QGLDPASQTKVVELLKDGSGR-GHRRGSRDAG---POSESDLESGPAEWSGDGSGSL 534

Query:   541 RRSQSGFKLRDALRRSSEMLVKKLQGGTPOEPPNPMKRASSLNLFLNKSVEEPTQGG 598
RRSQSGFKLRDALRRSSEMLV+KLQGG PQEP N RMRASSLNLFLNKSVEEP QPGG
Sbjct:   535 RRSQSGFKLRDALRRSSEMLVRKLQGGQPQP-NRMRASSLNLFLNKSVEEPVQPGG 591

```

Pedant information for DKFzphes3_4h6, frame 3

Report for DKFzphes3_4h6.3

```

[LENGTH] 622
[MW] 68934.82
[pI] 6.72
[MOBOL] TREMBL:AF055665_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus
Kinesin light chain 2 (Klc2) mRNA, complete cds. 0.0
[BLOCKS] BL00927C Trehalase proteins
[BLOCKS] BL01160I Kinesin light chain repeat proteins
[BLOCKS] BL01160H Kinesin light chain repeat proteins
[BLOCKS] BL01160C Kinesin light chain repeat proteins
[BLOCKS] BL01160P Kinesin light chain repeat proteins
[BLOCKS] BL01160E Kinesin light chain repeat proteins
[BLOCKS] BL01160D Kinesin light chain repeat proteins
[BLOCKS] BL01160C Kinesin light chain repeat proteins
[BLOCKS] BL01160H Kinesin light chain repeat proteins
[BLOCKS] BL01160A Kinesin light chain repeat proteins
[SUPFAM] tetra-tricopeptide repeat homology la-07
[PROSITE] RGD 1
[PROSITE] HYRISTYL 8
[PROSITE] KINESIN_LIGHT 2
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 5
[PROSITE] CK2_PHOSPHO_SITE 11
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 7
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] Kinesin light chain repeat
[KW] All_Alpha
[KW] LOW_COMPLEXITY 12.54 %
[KW] COILED_COIL 4.98 %

```

[illegible]

Prosite for DKF2phtes3_4h6.3

PS000001	449->453	ASN GLYCOSYLATION	PD00000001
PS000001	587->591	ASN GLYCOSYLATION	PD00000001
PS000004	425->429	CAMP_PHOSPHO_SITE	PD00000004
PS000004	505->509	CAMP_PHOSPHO_SITE	PD00000004
PS000004	554->558	CAMP_PHOSPHO_SITE	PD00000004
PS000004	578->582	CAMP_PHOSPHO_SITE	PD00000004
PS000004	616->620	CAMP_PHOSPHO_SITE	PD00000004
PS000005	30->33	PKC_PHOSPHO_SITE	PD00000005
PS000005	90->93	PKC_PHOSPHO_SITE	PD00000005
PS000005	415->418	PKC_PHOSPHO_SITE	PD00000005
PS000005	499->502	PKC_PHOSPHO_SITE	PD00000005
PS000005	507->510	PKC_PHOSPHO_SITE	PD00000005
PS000005	539->542	PKC_PHOSPHO_SITE	PD00000005
PS000005	615->618	PKC_PHOSPHO_SITE	PD00000005
PS000006	117->117	CK2_PHOSPHO_SITE	PD00000006
PS000006	151->155	CK2_PHOSPHO_SITE	PD00000006
PS000006	163->167	CK2_PHOSPHO_SITE	PD00000006
PS000006	232->232	CK2_PHOSPHO_SITE	PD00000006
PS000006	470->474	CK2_PHOSPHO_SITE	PD00000006
PS000006	517->518	CK2_PHOSPHO_SITE	PD00000006
PS000006	519->523	CK2_PHOSPHO_SITE	PD00000006
PS000006	521->525	CK2_PHOSPHO_SITE	PD00000006

PS00006	568->572	CK2_PHOSPHO_SITE	PDOC00006
PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	610->614	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	339->347	TYR_PHOSPHO_SITE	PDOC00007
PS00007	424->432	TYR_PHOSPHO_SITE	PDOC00007
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	482->488	MYRISTYL	PDOC00008
PS00008	598->604	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00009	292->296	AMIDATION	PDOC00009
PS00009	499->503	AMIDATION	PDOC00009
PS00016	502->505	RGD	PDOC00016
PS01160	223->265	KINESIN_LIGHT	PDOC00893
PS01160	265->307	KINESIN_LIGHT	PDOC00893

Pfam for DKF2phtes3_4h6.3

HHM_NAME Kinesin light chain repeat

HHM *RALEDREKTLGHDPDVA+MLNNLALVCRNKNYE+VENYYN*
 +ALE+EXT+GHDHPDVA+TMLN+LALV+R+QNKYE+*+N

Query 223 QALEDKTSGHDHPDVA+TMLN+LALV+R+QNKYE+EAHLLN 264

50.46 265 306 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain

Alignment to HHM consensus:
 Query *RALEDREKTLGHDPDVA+MLNNLALVCRNKNYE+VENYYN*
 AL+REKTLG DHP VA LNNLA+++ ++KY+E+E++ 306

dkfzphes3 265 DALATREKTLGKDHFAVAATLNNLAVLVGKRGKYKAEPLCK 306

Query 348 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain

Alignment to HHM consensus:
 HHM *RALEDREKTLGHDPDVA+MLNNLALVCRNKNYE+VENYYN*
 RALE+REK+LG HPDVA++L+NLAL+C+NQ+K EEVE YY+

Query 307 RALETRKVLGKFRHPDVA+QLSNLALLCQNGKAEVEYYR 348

39.10 349 390 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain

Alignment to HHM consensus:
 Query *RALEDREKTLGHDPDVA+MLNNLALVCRNKNYE+VENYYN*
 RALE+ LG D P+VA+ NNLA + Q+KY+++E +Y+

dkfzphes3 349 RALEIYATRLGPDHPVAKTKNNLASCYLKQGRYQDAETLYK 390

WO 01/12659

PCT/IB00/01496

DKFZphtes3_4o19

group: testes derived

DKFZphtes3_4o19 encodes a novel 1180 amino acid protein with weak similarity to human megakaryocyte stimulating factor and human mucin.

The novel protein contains a cytochrome c family heme-binding site signature.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to megakaryocyte stimulating factor and mucin

complete cDNA, complete cds, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 3767 bp

Poly A stretch at pos. 3757, polyadenylation signal at pos. 3737

```

1 GGCTAGGTTT AGCTTCAGGG GCAGCCAGG GCAGTGTTC TGCAATTGTC
51 ATGGATGAAA GGCCTAAGGC TGCCCTCTCT TGCAGGCTGG CTCTTGAGAT
101 TGCACCTTCT TCTCCTGCTA CTCTCCAAA TCTATGACCC TTCAAGGCAG
151 AGCTGACCTG TCGGTAAATC AAGGCAATGC AGCCGGCCCG CTAGCTACAG
201 TTACAGAGGC ATGTGTACGC CAGTGGGGGG TGCACTCTCC AGCCCCGCT
251 CACCCAGCTC TCTGGACAA AATGGAGAAA GCGCTCCAC AGCCCCAGCA
301 CGAGGGCCTC AAGTCCAAGG AGCATCTTCC GCAACAGCCT GCCGAAGGCA
351 AGACGGCGTC CCGCCGCGTC CCAGCGCTTC GGGCTGTGTT AGAGAGCCAG
401 GCTTTCAGAA ACATCTGGT AGACGAGATG GACATGATC AGGCCCTGTC
451 AGCCAGCGTC ATCCAAGCCA ATGGAGGGGG CTATTGGCTC CGGCAAGGCG
501 TGATTTCCTA GATGATGGCG GCCAAGGCCA TCCAGAGGC CTGGCGGCGC
551 TTCAACAAGA GACACATCTT TCACTCCAGC AAGTGGTTGG TAAAGAAAAC
601 GAGGGCGGAG GAGGGGGACA TACCTATCA CCCCACAGC CAGGTGGCT
651 TCCAGCATCC GGAAGAGAAC CGCTTCTGT CCCCAGCAT CATGCTGAAC
701 AAGGAGACCC AGTTCCCTTC CTGTGACAA CTGGTCTCTT GGAGACCCCA
751 GTCTGCTCCC CTCTGCGAG CCCCAGCAGC TCAGGGTACC CCAGAGCCCT
801 GTGTGCAAGG TCTCATGTCT GGCAGATGCC GGGGGTGGC CTCTCGCCA
851 CACGAGAGGG TGCATCAG ATTTCTCTGC CCAGTGAATT TGGAGCCAAA
901 ATGCCAGCCA TGCTGTGTA CCAGAACCAT CAGAAGCACC TGCTCTGCTC
951 ACATAGAGGG TGACTCACTG AAGACCAAC GTGTAGTGC CCGGACCAAC
1001 AAGGCCAGGG CTCGGAGAG ACCATTGTCC AGAAGGTATG ACCAGGCACT
1051 TACGAGACCA TCCAGAGCC AAGCCAGGG CCTGTGAAG GCGAGAGCCC
1101 CCAAAGCCCC CTTCAGATA TGTCAGGGCG CCATGATCAC CAAGACTCTA
1151 TCCAGACAT ATCCAGTGT CTCTGTGACC CTGCCACAGA CATATCCAGC
1201 GTCCAGCATG ACCACACCC CACCCAGAGC TAGCCAGTT CCCAAGTAA
1251 CAATATCAA GACCCACCC CAGATGTAT CCGGGCCAC AGTGACCAA
1301 ACTGCACTC ACACATGCC CATGCCACA ATGACCAAGA TCCAGGTACA
1351 CCCCAGAGCC TCCAGAACTG GCACCCACG GCAGACATGC CCTGCAACCA
1401 TCACGGCAAA GAACCGACTT CAGGTTTCCC TTCTGGCTTC CATCATGAAG
1451 AGCTTCCGCC AGGTATGCC GGGGCTGGG ATGGCAAGA CCCCACCCA
1501 GATGCCACCG GTACACCCC CAGCCAAAAA CCAATGCAA ACATGTCTGT
1551 CAGGCCAAT GTCCAAGACT TCATCCAGA GGAGCCAGT TGGGTGACC
1601 AAGCCCTCAC CCGAGACCG CCTGCCAGCC ATGATAACCA AGACCCAGC
1651 CAGTTACGC TGGGTGCCA CATCTCAA GACTCTGT CTGGCTCTC
1701 CAACAGTGGC AATGTCAA GCTCCACCC AAGTGGGT AGCAGCCGGA
1751 ACTCCAAACA CTTGAGGCT CACTCATGAG AACCCACCA AGGCCAAGGC
1801 CACCGTAAAT GTGAAGCAG CTGCAAAAGT GGTAAAGCC TCATCCCCCT
1851 CTTATTTCG TGAGGGAG ATCAGTGCC TGCTCAAC ACATCCGGA
1901 ACTGGGTTC CAGGGCTG AGCTGAGTT CTTTGAAG CCGAAGAAAT
1951 CAAGACTGGC ACCCAGAAC AGGCCAAGC AGCATGGCA TTTAAGACCA
2001 GTGTGGAGT GGAATGGCT GGGGTCCAT CCTGGACAAA AGTTGCTGAG
2051 GAAGGGGACA AGCCACCTCA CGTGTATGT CCTGTAGACA TGGCTGTAC
2101 CTGCCCCCG GAGACCTG CTGCCCAT GACCAATG TCATCCAGA
2151 GACATCCACC CTGCTGTCC CAGAGACCA TGCCGCCCC GCTGACCAAG
2201 GCTCATCTC AGGACATCT GCCCACTGAG CTGACCAAGA CCCCATCCCT
2251 GGCCATCTG GACACTGTC TGAGCAAGAT GCATTCCAG ACACATCTGG
2301 CCACAGGTGC GTGAAGGTC CAGTCCAG CCGCTTACG CACCTCTGT
2351 ACCAGAGCG ATCTCGGG GCACCCGATC ACAGACATA CCGCTGCTT
2401 CATCCAGCG CACCAAGCT CTGATCTCAG CAGCAACCC CACTCCAGG
2451 TGCTCTAAC AGGTCCAA GTGTCCAACC ACCTCTGCA GCGCTCGGT
2501 GGCCTCAGC CCCCACCTG GGCACAGCA GAGGACAGC AGACCCAGCC
2551 AGACCCAC GACACCTG CCGGAGAGC CACTCAGG GACCATGCC
2601 CGCAGCGCT TGAGTCCAG GGTATGCTG TGCCGCGAT GGCACCCACC

```

```
>TREMBL:HSU70136.1 product: "megakaryocyte stimulating factor"; Human
megakaryocyte stimulating factor mRNA, complete cds.
Length = 1,404
```

HSP9:

Score = 242 (36.3 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16
Identities = 145/546 (26%), Positives = 198/546 (36%)

[illegible]

Score = 205 (30.8 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12
Identities = 146/565 (25%), Positives = 209/565 (36%)

Query:	281	TKRSVANTHKKAPATPEPLSRRYQDTPSPRQATGCVRLAE--TPKAPQICQPMGITK	338
		K K A P T P A T P + P K P T P P + T	
Sbjct:	597	TRKPAPATPEKPAFTPTK-----ETAPFTPKLITPTPEKLAPITPEKPAFTPEELAPIT	652
Query:	339	LLQTPYVSVVLTLPQATPSTVITTPPKTSVP--PVRTIITKPAQNYGPTVTVL--TAPHTCP	396
		P T P T P T P T P T P T P T P T P T P T P T P T P T P T P T	
Sbjct:	653	PEEPTPTPEEPAPATPKAAAPITPEKPAFTPEKPAFTPEKPAFTPEKPAFTPEKPAFT	709
Query:	397	M-----PTMTKIQVHTASRTGTPTQTCPATITAKNRNPQSVLLASIMESLQVCVCPAMAKT	453
		P T K + P T + P P P P T + S + K P G A T	
Sbjct:	710	KGTAPTEKLA-EAPATPTPKPAKELAPIT-----TKEPTISSD--KFPKGTAPIT--	761
Query:	454	PPQHPVPTTPAKNPLQSLTASHTKSQSRVGVTKPSQSPRLAMITITPAQLRASVAT	513
		P + P TTP K P T T T + R P + P K T P S	
Sbjct:	762	PKPEAP--TPT--KEPAFTPKGTATPPTPKPAKELAPIT--TNGPITSVSDK	818
Query:	514	ILTKLISLSTPYNANVPKAPVVAAGATPTSGSIEHNPKKAKATVNV-----QKAAKFA	569
		T P T A P A P T E P P + V + K + K +	
Sbjct:	819	ATPTPKETAPATPTPTPTPKPKA--PTTP-----ETTPPTVSSTPTTKEPTIHK	872
Query:	570	---SSPSYLAEGKILCLAQHPHQTGPRAAEELFLAEKGTIKQQAQKATKSTSVAG	626
		S S P A E + L + L G V F P + P + T T X K T T + *	

WO 01/12659

PCT/IB00/01496

```
2651 GGCCATTCCA CATGCAACGT TCAGTCCTGG GGAGACAACG GAGCCACACG
2701 TGCCAGGCEA TCAATGCCCG GCGAGGCGGT GCCCTGCCAG GAGCCACACG
2751 GCGCCGGGGA CGCTGGTGTG GTTGCTGGCT AATCTGGGAA CGCGCATGG
2801 GAGCCAGCEA GGGGTGCTGC GTCTGGGAC ACCTGGGCEA ACAAGGCGGT
2851 GGTGCTCTCC AGGCGGTCCG GGGAGCCCAT GGTGTCCATG CAGGCTGCAG
2901 AGGAGATCCG CATCTTCGCA GTGATCACTA TCCAGGCGGG CGTCCGTGGC
2951 TACCTGGGCG GTGCGAGGAT CCGGCTGTGG CACGGGGGGG CCATGCTCAT
3001 CCAAGCTACT TGGCGCGGCT ACCGTGTGGG GCGGAACCTG GCACACCTCT
3051 GCAGAGCCAC CACGACCATC CAGTCTGCCT GGGCGGGCTA CAGCACCCGC
3101 CGGGACCAAG CCGGCACTG GCAGATGCTC CACCCCTCA CTTGGGTGGA
3151 GCTGGGCGCG CGCGCGGGG TCATGCTGTA CGAAGCTGG TTCCAGGATG
3201 GCAGAGCCAG GACAGTATCT GACCATGCTT GCTTCCAGTC CTGCGAGGCA
3251 CAGGCTTGCA GCGTCTGCCA CTCCTGAGC TCCAGGATCG GGAGCCGCC
3301 CAGGCTGTGT ATGCTAGTGG GCTCCAGCCC TCGCACCTGT CATACTGTGT
3351 GAGCCACACA GCCCACCCTG GTGGTGAGG GCATGGGCEA GGCACATGAC
3401 GCGCCGGGGG CAGTGTCTTG GCGCTCGGCT TACGAGCTGG CTGGCTCTAG
3451 TCCGAGGCGC CCGCATGCCC AGGACAAGC GGCACAGCCG ATCCAGTCCG
3501 CCTGGAGGGG CTTTAAGATC CGCCAGCAGA TGAGGAGCEA GCAATGGCA
3551 GCGAGATAG TTCAAGCCAC CTGGCGAGGC CACCATACCC GGAGCTGTCT
3601 GAGGACACA GAGGCGCTCT TGGAGCAGG AGACCCCTGG GCGAGCTCAC
3651 GGCACATGCA TTGGCCTGGC ATCTAGGACC CTGGCTCCCT GCAGTGGGGA
3701 CTTCGTGGGA GGCACCTCAT GCTCTCTGGG TCTAATGAAT AAAGTCTCC
3751 ACAGCCTAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3673 bp: peptide length: 1180
Category: similarity to known protein

```
1 MTIQGRADLS GNQNAAGRL ATVHEPVVTO NAVHPPAPAH PSLLDKMEKA
51 PPQPQHEGLK SKEHLPOOPA EGKTASRRVP RLRAVVSQA FKNILVDEMO
101 MHUARAATLI QANWRGYWLR QKLISQMAA KAIDANRRF NRRHILHSSK
151 SLVKRTAEI GDIPIVIAPOQ VRFQHPEDRI LLSPPIMVWK ETQFPGCDL
201 VLRCPOSSPL LQPPAAGTPT EPCVQPHAA RVRLAFLPH OTVIRFPCL
251 VSLDAKCPCL LLTRTIRSTC LVHIEGDSVK TKRVSARTNK ARAPETPLSR
301 RYQAVTRPS RAQTQGVVKA ETPKAPFQIC PGPHITRILL QTPVVSUTL
351 PQTYPASTNT TTPKTSFVP RVIIITKPAQ NYGPPVTIKT APHICDMPHT
401 TKIQVHTAS RTCTPRQTCF ATITAKNRPO VSLIASINKS LPQVCPGPM
451 AKTPPQNHVP TTPAKNPLQT CLSATHSKTS SQSPVGVTK PSQTRLPLM
501 ITRTPAQLRS VATILKTLCL ASPTVAHVRA PPQVAVAGT PNTSGSINEN
551 PPSKATVNV KQAAKVVEAS SPSYLAEKI RCLAQPHPT GVPRAAEELP
601 LEAEKIKTCT OKQARTDMAF KTSVAVENAG APSWTKVAEE GQKPPHYVVP
651 VMAVTLPRG OLAAPLTHAS SQRHPPCLSQ RPLAAPTKA SSQGHLPTEL
701 TKTPSLAHLQ TCLSRHMSQT HLATGAVKVO SQAPLATCLT KTQSRGQFIT
751 DITTCLEPAM QAADLSNTH SQVLLTGVSK SHHACORLGG LSAPPMWKE
801 DRGTQDPHG HVQRTITGCG CPKACVQCG HLPVPMAPTG HSTCNVESWG
851 DNGATRAQPS MPQAVPCQE DTCPADAGVV GGQSNWAME PARGAASWD
901 WRNKAVVPR RSQEPVSMQ AAETIRILAV ITIQAGVRY LARRIRLWH
951 RGAMVIGATM RGVNVRNLA HLCRATTITQ SHWRGVSTRR QDARHQLHM
1001 PWTVEIGSR NGVMSDSWF GQCRATVSD HRCFQSCQAH ACEVCHSLSS
1051 RIGSPSPVVM LVGSSPRTH TCGRTQPTRV VQMGQGTGEG PCAYSWASAY
1101 OLAALSPROP HRQDRAATAI QSANRGFKIR QMRQOQMAA KIVQATWRGH
1151 HTRSLKNTK ALLGPADPSA SSRHMHWGI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DNF2phtes3_4019, frame 2

TREMBL:HSU70136.1 product: "megakaryocyte stimulating factor"; Human
megakaryocyte stimulating factor mRNA, complete cds., N = 2, Score =
242, E = 9.6e-16

Sbjct: 873 PDESTPELSAEPFKALENSPKPGVP--TTKPAATKPEMTTAKDKTTERDLATTPET 930

Query: 627 EMAGAPSWTK-VAEEGDKPHVYVVDNAVTLPRGQLAAPLTNASSQRHPPCLSORPLAA 685
A AP TK A +K +T Q+ + T + L LA

Sbjct: 931 TTA-APKMKETATITTEKT-----TESKITATTQVSTTTQDTPFKITTLKTTTLAP 983

Query: 686 PLTKASSOGHLPTLTKTSPSLAHLDTCLSKMHSOTHLATGAVKVQS-----QAPLATCLT 740
+T + + TEV P +T K + AT K Q + P +T

Sbjct: 984 KVT-TTKKITITTEINMKPE----ETAKPKORATNSKAT-TPKPKPTKAPKPTSTKKP 1037

Query: 741 KTQSR-GQPTITDIT----TCLIPAHQAADLSNTHSOVLGTGSKVSNHACQRLGGLSAPP 795
KT R +P T T +P Q + + H + S

Sbjct: 1038 KTMFRVRKPKTTTPRKMTSTHPELNPTSRIAEAMLQTTTRNQTTPHSLVEVNPXSEDA 1097

Query: 796 W-AKPEDROTOPQPHGVPGKTTGGPCPAACEVOGMLVPPHAPTHGSHCN 845
A+ E +PH +P T P OG+++ PH + CN

Sbjct: 1098 GGAEGETPHMLLAPHVHPEVTPMDYLPVRVP-QGIIINHLSDETNIEN 1147

Score = 198 (29.7 bits), Expect = 2.3e-11, Sum P(2) = 2.3e-11
Identities = 142/513 (27%), Positives = 200/513 (38%)

Query: 204 RPQSSPLLQPPAAQGTPEPCVOGPHAAVRGLAFLPHQTVTIRFPCPVSLDAKQCCLLT 263
R + P +PP G + H V+ + +P L

Sbjct: 207 RTKKKPTPKPPVDEAGSGLONGDFKVTTPDSTTHQHNKSVSPKITTAKPINPRPSLPP 266

Query: 264 R--TIRSTCLVHIEGDSVTKRVSARTNKARAP---ETPLSRRYQAVTRPSR---AQTQ 315
T + T L + +V+TK + TNK + E S + Q+ + S A T

Sbjct: 267 NSDTSKETSITVNMKETTETKETT-TNKQTSIDGKETTSAKETOSIEKTSAKDLAPTS 325

Query: 316 GPVKAETPKAPFOICPGPMITKLLQTYPVVSVTLPTQYPASTMTTTPKTSVPVKVII 375
+ TPKA GP +T T + P T P+ PAST TP + P +

Sbjct: 326 KVLAKPTPKAE-TTNGPALT-TPKEPTP---TPKPE-PAST---TPKEPTTITKSAP 375

Query: 376 KTPAQVYGPVTVTAPHTC---PMTMKIQVHTASRTGTPROT-PATITAKNRQVS 432
TP + P PT TK+AP T P PT TK + PT + P T PA T K+ P

Sbjct: 376 TTPKE--PAPTTTSAPTTPEKEPAPTTTK-EPAPTTPEKAPPTTKEPAPTTTSAPTT 432

Query: 433 ---LLASINKSLPQVCPGAMAKTPQKHVPTTPAKNPLQCLSATNSKTSQSSQSPVGV 489
+ K P PA TP + P TTP K P T + T + +P

Sbjct: 433 KEPAPTTPKKAPPTTPEKEPAPT-TPKEPTP-TTP-KEPAPTTKEPAPT-TPKEPAPTAPK 488

Query: 490 KPSPOT-RLPAMIT-KTPAQLRSVA---TILK---TCLASPTIVANVKAPQVAVAGT 540
KP+P T + PA T K PA + T K T +PT AP A T

Sbjct: 489 KPAPTTPEKAPPTTPEKAPTTTKEPSPTTPEKAPTTTNSAPTTTKEPAPTTTSAPTT 548

Query: 541 PNT-SGSIHENP----PKAKATVNVQAAKV-KASSPSYLAEGKIRCLAQPHGTGVPR 594
P S + + P PK A K+ A K +P+ E +P P P+

Sbjct: 549 PKEPSPTTKEPAPTTPEKAPTTTKEKAPTTTPEKAPTTTPEKAPTTTKEKAPTTT 606

Query: 595 AAAPLEAEKIKTGQKQARTOMAKTSVAVEMAGAPSWTK-VAEEGDKPHVYVVDN 653
A P + + T K+ K + AP+ +A + P P +

Sbjct: 607 EPA---PTTPEKAPTTPKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEP 664

Query: 654 AVTLPRGQLAAPLTNASSQRHP-PCLSORPLAAPLTKASSOGHLPTLTKTSPSLAHLDT 712
A T P+ AAP T + P P + P AP T P E T T

Sbjct: 665 APITPKA--AAPNT----PEKAPTTPEKAPTTTKEKAPTTTPEKAPTTTPEKAPTTT 716

Query: 713 LSK 715
L +

Sbjct: 717 LKE 719

Score = 108 (16.2 bits), Expect = 4.3e-02, Sum P(2) = 4.3e-02
Identities = 60/214 (28%), Positives = 85/214 (39%)

Query: 265 TIRSTCLVHIEGDSVTKRVSAR-TNKA--RAPETP-LRRYQAVTRPSRAQTGQPVKA 320
T + +H D T +SA T KA +P+ P + A T+P T

Sbjct: 862 TTKEPTTIHKSPE-STEPLSAEPTKALENSPKPGVPTTKTPAATKPEMTTAKDKTT 920

Query: 321 ETP--KAPFOICPGPMITK-TLLQTYPVVSVTLPTQYPASTMTTTPKTSVPVKVII 377
E P P +PK T T + T TTT T+D E+T KT

Sbjct: 921 ERDLRTTPETTTAAPKMKETATITTEKTESKITATTQVSTTTQD-TTFF-KITTLKT 978

Query: 378 PAQVYGPVTVK---TAPHTCPMPTHT-KIQVHTASRTGTPROTCPATITAKNRQVSL 433
+ P T TK T P T K + T S+ TP+ P A +P +

Sbjct: 979 TT-LAPKVTITTKITITTEINMKPEETAKPKORATNSKATTPKPKPTK--APKPTSTK 1035

Query: 434 LASINKSL--PQVCPGPA-NAKTPQKHVPTTTPARNPLOT 470
H + P+ P P K T P+++P + A+ LQT

Sbjct: 1036 KPTMFRVRKPKTTTPRKMTSTHPELNPTSRIAEAMLQTT 1075

Score = 56 (8.4 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12

Identities = 17/60 (28%), Positives = 22/60 (36%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAEGKTASRRVF 80
T EP T P P PS E AP P+ + K+ P P E + + P
Sbjct: 533 TTKEPAPTTTKSAPTTKPEPTTTKEPAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEP 592

Score = 52 (7.8 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16
Identities = 17/59 (28%), Positives = 22/59 (37%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAE-GKTASRR 78
T EP T P P P+ E P P+ +KE P P E TA ++
Sbjct: 431 TPKEPAPTTKPKPAPTTKPEPAPTTKPEPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 489

Score = 51 (7.7 bits), Expect = 1.2e-15, Sum P(2) = 1.2e-15
Identities = 15/51 (29%), Positives = 19/51 (37%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAE 71
T EP T P P P+ + AP P+ + KE P P E
Sbjct: 416 TTKEPAPTTTKSAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPTTTKPE 466

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
Identities = 12/41 (29%), Positives = 17/41 (41%)

Query: 36 PAPANPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKTAS 76
P P P + P +P +KS P++PA T S
Sbjct: 350 PTPTTK--EPASTTKPEPTTTIKSAPTTKPEPAPTTKS 388

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
Identities = 15/57 (26%), Positives = 19/57 (33%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEG-LKSKEHLPOQPAEGKTASR 77
T EP T P P P+ E AP P+ +KE P T +
Sbjct: 377 TPKEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKKSAPTTK 433

Score = 46 (6.9 bits), Expect = 4.0e-15, Sum P(2) = 4.0e-15
Identities = 16/58 (27%), Positives = 22/58 (37%)

Query: 20 LATVHEPVVT---QWAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKT 74
L T EP T + A P P+ + P +P KS P++PA T
Sbjct: 344 LTPKPEPTTTKPEPASTTKPEPTTTIKSAPTTKPEPAPTTKSAPTTKPEPAPTTT 401

Score = 42 (6.3 bits), Expect = 1.0e-14, Sum P(2) = 1.0e-14
Identities = 15/60 (25%), Positives = 21/60 (35%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAEGKTASRRVF 80
T EP T P P P+ + AP P+ + KE P E + + P
Sbjct: 463 TPKEPAPTTKPEPAPTTKPEPAPTAPKKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 522

Score = 39 (5.9 bits), Expect = 2.1e-14, Sum P(2) = 2.1e-14
Identities = 15/55 (27%), Positives = 20/55 (36%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKTAS 76
T EP T P PA + + P +P KS ++PA T S
Sbjct: 494 TPKEPAPTT----PKEPAPTTKPEPTTKPEPAPTTKSAPTTKPEPAPTTKS 544

Pedant information for DKF2phtes3_4ol9, frame 2

Report for DKF2phtes3_4ol9.2

[LENGTH]	1180
[MW]	127693.40
[PI]	10.25
[HOMO]	SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2). 1e-08
[FUNCAT]	98 classification not yet clear-cut [S. cerevisiae, YJR151c] 6e-06
[FUNCAT]	30.01 organization of cell wall [S. cerevisiae, YI019c] 6e-06
[FUNCAT]	30.90 extracellular/secretion proteins [S. cerevisiae, YI019c] 6e-06
[FUNCAT]	01.05.01 carbohydrate utilisation [S. cerevisiae, YI019c] 6e-06
[BLOCKS]	BLO0412B Neuromodulin (GAP-43) proteins
[PROSITE]	CYTCHROME_C 1
[PROSITE]	MYRISTYL 12
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 8
[PROSITE]	PKC_PHOSPHO_SITE 25
[PROSITE]	ASN_GLYCOSYLATION 2
[KW]	Alpha beta
[KW]	LOW_COMPLEXITY 5.00 %

895

Prosite for DKFZphtes3_4o19.2

PS00001	542->546	ASN_GLYCOSYLATION	PDOC00001
PS00001	668->672	ASN_GLYCOSYLATION	PDOC00001
PS00004	282->286	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	278->281	PKC_PHOSPHO_SITE	PDOC00005
PS00005	281->284	PKC_PHOSPHO_SITE	PDOC00005
PS00005	285->288	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	424->427	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	610->613	PKC_PHOSPHO_SITE	PDOC00005
PS00005	671->674	PKC_PHOSPHO_SITE	PDOC00005
PS00005	679->682	PKC_PHOSPHO_SITE	PDOC00005
PS00005	900->903	PKC_PHOSPHO_SITE	PDOC00005
PS00005	959->962	PKC_PHOSPHO_SITE	PDOC00005
PS00005	987->990	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1015->1018	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1049->1052	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1065->1068	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1106->1109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1146->1149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1171->1174	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	988->992	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1003->1007	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1027->1031	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	591->597	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	777->783	MYRISTYL	PDOC00008
PS00008	853->859	MYRISTYL	PDOC00008
PS00008	878->884	MYRISTYL	PDOC00008
PS00008	882->888	MYRISTYL	PDOC00008
PS00008	1008->1014	MYRISTYL	PDOC00008
PS00008	1053->1059	MYRISTYL	PDOC00008
PS00008	1083->1089	MYRISTYL	PDOC00008
PS00190	1042->1048	CYTTOCHROME_C	PDOC00169

(No Pfam data available for DKFZphtes3_4o19.2)

WO 01/12659

PCT/IB00/01496

DKFiphtes3_50j4

group: testes derived

DKFiphtes3_50j4 encodes a novel 187 amino acid protein proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 1186 bp

Poly A stretch at pos. 1176, polyadenylation signal at pos. 1126

```
1  CACTGGGGCT CTGAAGCTCA GAGCTCACCC CTGAGATGGG CTCTCTAGG
51  CCTCTCTGGA TGAGGGAGCC ACCAGGACCC AGTCTGTGA TGCCTGCTCT
101 CCTCTCTACC AGCAGCTGDC GGGCCAGAGA CCAGGGCACC CCTGAAGTCC
151 AGCCGACGCC TGCAGAGGAC ACATGGGAGG GCAGGGGGCC TGGATCCGAG
201 CAGGAGAACCC CAGAGAGCCA GCCTCAGAAG AGGCCAGGCC CCTCAGCCAA
251 GGCCTCCCTC GTAGCTGAGG TCAAGGGCAG GGTCTGGGCC AGCGAACAGG
301 GCACCTTGAA TCCACGGGCT CAAGACCCCT TCCAGCTCTC GGTCTCTGGC
351 GTCTCTCTTA AGGAGGCTGC AATGTGTGTG CTCAGTGGCC TCACCCCTTT
401 CTACAAGGAG GGCAAGTTTG CTTCGAAGGA GTTGTTTAAA GGTCTTGGCC
451 GGCACCTCTC ACACCTTGCTG ACTCAGAAGA CCTCTCTGGG AAGGAGCGTG
501 AAGAAGAGAG CCCAGAACCT CATCAGGCAC TTCTTCCATG GCGGGGCCCG
551 GTCCGAGAGC GAACTGTACT GGCATGGCTT GTGTGGCCCC CAGAGATGAC
601 CACTCTCTGG CTGGGCAGGG CCCCCTCTCT CCCCAGATT CTAGCATGGG
651 TCATCTCTGG CCTCACCTGC TGATGCCAGG GGCATGCTCT TTTCTAGTC
701 CTCTCTCTTT CCAACCATAC TTGGCTTTGG GGATGACCCG AGACACCCCG
751 TGAATCCAGG TCAAGAGTCA GGCACCTTTT TTTCTGCTT GCAGAGCTTA
801 TAGACCTTTC TCAGAGCGGT CCTCATGGCT GGTTTCTG TGACACATGT
851 CGAGGACAGA AGGTGGAGGG TGGTGGAGCT GCTCTGGAA GAAGGGGAAG
901 GAAGAGTGGC CCTCTCCCGA GTTCTAAGTC AGGATGAGGC CCACCTGTCC
951 AAGGTATCGG AACTTACCCA GGGGAGCCTC AGATCTCCGA CCCACTGCCG
1001 CATCTATTAC GATGCGAGCT TCCAGCCTTG CCAAGTCCAG AGCTGTGGCA
1051 GAGGAGAGGC AGCCAGGCCG TGTCTCTGCT CAGCTCTCTG TCAGGAAGGC
1101 CAGGCCCTGAC AGATGTTTGG GAGAGGAATA AAGTTGTGTT GTTGTGGGGC
1151 ATGCAGCGGT GCACACAGCC CTTTCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 36 bp to 596 bp; peptide length: 187
Category: putative protein

```
1  MGSPPRPQMR EPPGPSAVMP ALPSTSTCPP RDQCTPEVOP TPAKDTWRGK
51  RFRSQENPFE SPPQNRPPFP AKFSVVAEVK GSVSASEQOT LMPFAGOPFQ
101 LSAQGVLEKE ANPVVVKLIT PTYKEGRFAS KELEKGFASH LSKLLTQKTS
151 PGRSVKEEAO NLIRHFFHGR ARCESEADWH GLCGPQR
```

BLASTP hits

Entry MMU92455_1 from database TREMBL:

WO 01/12659

PCT/IB00/01496

product: "WW domain binding protein 7"; Mus musculus WW domain binding protein 7 mRNA, partial cds.
Score = 134, P = 6.9e-08, identities = 45/125, positives = 56/125

Alert BLASTP hits for DKFZphtes3_50j4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50j4, frame 3

Report for DKFZphtes3_50j4.3

[LENGTH] 187
[MW] 20353.06
[pI] 9.76
[PROSITE] MYRISTYL 1
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 6
[PROSITE] PKC_PHOSPHO_SITE 6
[KW] All_Alpha
[KW] LOW_COMPLEXITY 8.56 %

SEQ MGSPRRPPGMRPPGSAVMPALPSTSTCPRDQGTPEVQPTAKDTWKGRPRSQENPE
SEG xxx
PRD ccc
SEQ SQPQRPRPSAKPSVVAEVKGSVSASEOGLNPTAQDPFOLSAPGVSLKEAANVVKCLT
SEG
PRD ccc
SEQ PFYKEGKFASKELFKGFARHLSHLLTQKTSFGASVKEEAQNLIRHFTGPARCEADMH
SEG
PRD ccc
SEQ GLCGPQR
SEG
PRD ccccccc

Prosite for DKFZphtes3_50j4.3

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	70->73	PKC_PHOSPHO_SITE	PDOC00005
PS00005	107->110	PKC_PHOSPHO_SITE	PDOC00005
PS00005	146->149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00009	48->52	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_50j4.3)

WO 01/12659

PCT/IB00/01496

DKFZphtes3_50n06

group: testes derived

DKFZphtes3_50n06 encodes a novel 186 amino acid protein without similarity to known proteins.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1095 bp

Poly A stretch at pos. 1085, polyadenylation signal at pos. 1061

```
1 CAAGACCCCTC GGAGCCAAGA AACAACTG AGTTCCAGAT TTCGGAAGGT
51 TCACGAGTGT TGCCGACACG CCTCCCAAC TGCAGACATC CTCCTTGGAG
101 GACCTGCTGT GCTCACAATG CCCCCTGTCC AGCAGAGACG ACACCTTCCC
151 GGCTCTGTCA GCGCCCTCC AGCAACCTT CAAGGCTTTC CTCATGTCCC
201 CAGAGCCACA TAGCCACCGA GGCACCGACA GGAAGCTGTC CCGCTCCTGT
251 AGCCCTCTTC AAGACTCACT GGTGGACAAG ACCCTGCTGG AGCCCAAGGA
301 GATGCTCCGG CCTAAGAAGG TGTGTTTCTC GGAGAGCAGC CTGCCACCGG
351 GGACAGGAGC CAGAGAGAGC TACTACCTCA ATGAGATCCA GAGCTTCGGG
401 GCGCCCGAGA AGGACGCGCG CCGTGGTGGC GAGATCGCTT TCCAGCTGGA
451 CCGCCCGCATC CTGGCCTACG TGTTCCTGGG CGTACGCGCG CTCTACGGCT
501 TCACGCTGGC CAACATCCCG GAGAAGATCG AGCAGACCTC CACCAAGTCT
551 CTGGAGCGCT CCGTGACGCA GAGGAAGCTG CGGAGCTGTA CGCAGCGCTA
601 CTGGGCTCTG AGCCGCGCGC TGGAGAAGCT CGGCTACAGC CGCGACGTGC
651 ACCCGGCGTT CAGCGAGTTC CTCATCAACA CCTACGGAAT CCTGAAGCAG
701 CGGCCCGACC TGCGCGCCAA CCCCCTGAC AGCAGCCCGG CCGCGCTCGG
751 CAAGCTGTGC ATCGAGCTGG TGCCGCCCA GTTCTGGGG GACTCCCTGC
801 TGCTGTCTAA CTGCTCTGCG GAGCTCTCCA AGGAGGACGG CAAGCCCTTC
851 TTCGCTGTGT GAGCGGCCCG GCGCCCGCGC CTTGCTCTGC AGTAAACGGC
901 TTTGTTCCAA CCGCGGCGCG CGGTGCTCTC TGGCGCTCCC CCGGAGGGGG
951 AAGGGCGCGC GTCCCGCGCG CCGAGGCTCA GAGAAAGCCC CGCTCCACC
1001 GGTGCTGGGC CCGACCGCA GCGCCGCGCT GCGCGACCT GCGGAGTCTC
1051 TCTCACCCCT CATTAAATC ATCGCTTTCG TTGTCAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 302 bp to 859 bp: peptide length: 186

Category: putative protein

Classification: no clue

```
1 MVRPKKVCFS ESSLPDGRD RRSYYLNEIQ SFAGAEDAR VVGEIAFOLD
51 RRILATVFPQ VTRLYGPTVA NIPEKIEQTS TNSLDGSVDE RKLRELTORY
101 LALSARLEKL GTSKDVNPAF SEFLINVTGI LKQRPOLRAH PLHSSPAALR
151 KLVIDVVPFK FLGDSILLLN CLCELSREGG KPLFAV
```

BLASTP hits

No BLASTP hits available

WO 01/12659

PCT/IB00/01496

Alert BLASTP hits for DKFZphtes3_50n06, frame 2
No Alert BLASTP hits found

Pedant information for DKFZphtes3_50n06, frame 2

Report for DKFZphtes3_50n06.2

[LENGTH] 186
[MW] 21049.39
[pI] 9.28
[KW] All_Alpha
[KW] LOW_COMPLEXITY 5.38 %

SEQ MVRPKKVCFSSESLPTGDRTRRSYVLWEIQSFAGAEDARVVGELAFQDORRILAYVFP
SEG
PRD ccc
SEQ VTRLYGFTVANIPKIEQTSTKSLDGSVDERKRLRELQRYLALSARLEKLGYSRDVHPAF
SEG
PRD ccc
SEQ SEFLINTYGIKQRPDLRANPLHSSPAALRLVIDVPPKFLGDSLLINCLCELSKEDG
SEG
PRD hhhhhcc
SEQ KPLFAN
SEG
PRD ccccc

(No Prosite data available for DKFZphtes3_50n06.2)

(No Pfam data available for DKFZphtes3_50n06.2)

WO 01/12659

PCT/IB00/01496

DKF2phes3_50n23

group: testes derived

DKF2phes3_50n23 encodes a novel 499 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits
(from other testis libraries) testis specific cDNA?

Sequenced by DKF2

Locus: unknown

Insert length: 1907 bp

Poly A stretch at pos. 1897. polyadenylation signal at pos. 1872

```
1 GGGCACCAGC CACTTTCNC CATGACTGTG CGCTCGAGGG TCGCAGATGT
51 GTTCGGCAGC AAGGACACTG AGAGCCTTGA GCCTGTGCTT TTACCCCTTAG
101 TAGATCGCAG GTTTCCTAAG AATGGGAAA GACCGGTGGC AGAAAGCTTA
151 GGGCACAAG ACAAGACCA GGAGGACTAC TTCCAGAGGG GAGGACTCCA
201 AATTAACTTC CACTGTAGCA AGCAGCTGTC TCTAGAGGC TCAGGCGAGG
251 TGACCTCTGA GAGCAAGAG CAGCCTGAGG AGGAGGAATT CGGCCGGGAG
301 ATGCGGAGGC AGCTGTGGCT GGAGGAGGAG GAGATGTGGC AGCAGCGGCA
351 GAGGAAGTGG GCCCTGCTGG AGCAGGAGCA TCAGGAGAAG CTGCGGCAGT
401 GGAATCTGGA AGACCTGGCC AGGAGGACAC AGCGAGATG GGTCCAGCTA
451 GAAAGAGAGC AGGAGGCTCC AGCGAGAGAG CCAGAGGAGC TAGGGGAGGA
501 TGTGAGAGGG AGGATCTTCA CACCCACCAG TCGATGGAGG GACTTGGAGA
551 AGCGAGAGCT ATCATTAGTG CTTGCCCCAA GCCGAGCCCA ATCTGCTCAC
601 CAAAGCAGGA GGCCACACTT GCCCATGTCT CTAATAGCC AGCAGCCTGC
651 CTTGGGAAG CAGAGACTTA TGACTTCAGT GGACZTTTACC TACAGACCCN
701 GGACCCGCCG AGTTCGCACA AAGCCCAAGA AATCTGCTTC CTTTCTGTCT
751 ACTGGGACAT CCATCCGAGG GCTGACCTGG CCCTCTTTGC AGATATCCCC
801 TGCAATATT AAGAAGAGG TGTACCATAT GGACATGGAG GCCCAGAGGA
851 AGAATCTGGA CTCTCTGAGT GAGGAGCTGT AGTTGAGGCT GCCCCACTAC
901 CTGCGAGGCA AAGCACTGGA GCTCACCACC ACCACATGG AGCTGGGCGC
951 GCTCAGGCTG CAGTACCTGT GCCATAAGTA CATCTTCTAT AGACGCCTCC
1001 AGAGCCTCCG CCAAGAGGCG ATCAACCATG TACAATCTAT GAAGAAGACG
1051 GAGGCTTCTT ACAGGCCCA GAACTCTTAC ATCTTCTCTG AAACATTGGA
1101 CCGCTCGCAG AGTCTCAGGC TGCAGGCTCG GACGGAACAG CAGAAGGGGC
1151 TGGAGGAGAA GCACCAGAGG TGCTTGAGCA GCATGTGTAC CATGTTCCCC
1201 AAGCTCCAGC TGGAGTGGAA GGTTCACCTG AACATCTCTG AGGTCACTTC
1251 GCAAGGCCA AAGAATGCA AGTTGCTCTG AGCTCTCCGC CGGCACTCC
1301 GCCCAGTGG CCCCACCTAC AAGCAGCCCT TTCTGTCTAG GCACCGGGCA
1351 TGTGTGCCCC TGCAGATGGC CCGCCAACAG GGAAGCAGA TGGAGGCTGT
1401 CTGGAAGACC GAGGTGGCCT CTCTCAGTGA CCAATAGAA AAAAGAACCC
1451 CTGCAAGCTT TCCCGGGAC CAGCTGAGGG GAGACCCAGA TATTCCTGGG
1501 CTGTTGACAC TGGAGCTGTA GTCTCTCTGC CACAAGGCC TGAATTCCT
1551 GAAGGCCAGC TAAGCGCTTC AGCGAACCAA AGGAAGGAAT GCCAGGAACC
1601 TACAAATGAA TCCGCTTAGC TTGTTCAAAA AAAGTCAAGG GAGTCACTCC
1651 CTGGAGCCCA AATAGGACAG AAGGATCAAG ACAGCCCGAG TCTCCACTGC
1701 ATCCCTCAGC CAGTGATTCT CAACCTTCTG AGGGACGAAA ACCCAGAGAG
1751 AACTTGGTCA AATGACAGGT TCCAGCTGGT TGCTTTTAAA GAAACCTCTT
1801 GGGGGTTGCT GAGTACTCCT AGAATTTTGA GAAACACTGC TTCCCTCTCTG
1851 CAGTCCCAA ACTCTACATT TTAATAAAT AGAGTTGCT TTATTTTAA
1901 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 1518 bp; peptide length: 499
Category: similarity to known protein
Classification: no clue

1 MIVRSRVADV FGSKOTESLE PVLLPLVDORR FPKKWERPVA ESLGKKKKDQ
51 EDYFQGGGLQ IKFHCSKQLS LESSRQVTSE SQEEPWEEEF GREHRRQML
101 EEEEMWQOQK RNWALLEQEH QKLRQWLE DLAREQORRW VOLEKQESP
151 RREPEQLGED VERRIFTPPTS RWRDLEKAL SLVPAPSRTO SAHOSRRPHL
201 PMSPTQOOPA LGKQRPNSSV EFTYRPRTRR VPTKPKRSAS PPVTGTSIRR
251 LTMPSLQI3P AMIKKVVYHN DNEAQRNLUQ LLSSESLRL PHYLRKALE
301 LTTTWMELGA LRQLVCHKY IFYRRLQSLR QGAINHVQIM KTEASYHAQ
351 NLVIFLENIQ RLQSLRLQAN TDQKQGLEK HRECLSMVT MFPKLQLENN
401 VHLNIPVETS PKPKCKLPA ASPKIRPSG PTYKQFLSR HRACVPLQNA
451 RQQKQMEAV WKTEVASSY ATEKTPASL PRDQLRGHPD IPRLLTLQV

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZp385t3.50n23, frame 1

PIR:S28589 trichohyalin - rabbit, N = 1, Score = 134, P = 5.3e-05

TREMBL:AF132479.1 product: "Ese2L protein"; Mus musculus Ese2L
protein mRNA, complete cds., N = 1, Score = 130, P = 0.00017

>PIR:S28589 trichohyalin - rabbit
Length = 1,407

HSPs:

Score = 134 (20.1 bits), Expect = 5.3e-05, P = 5.3e-05
Identities = 88/354 (24%), Positives = 154/354 (43%)

Query: 29 RRFPKKWERPVAESLGHKKKKDQEDYFQGGGLQIKFHCSKQLSLESSRQVTSESQEEPWEE 87
R++ K +R + L + ++E ++ G + F +OL +++ E +EE +
Sbjct: 165 RQYRDKERLQRLQDELEERRAEEQLRRRGRDAEEFTTEELRRRQQLKRELREEQ 224
Query: 88 EEFGRMRRQMLLEEEEMWQOQKRNWALLEQEHQKLRQWLEDLAREQORRWVOLEKQ 147
RE + L+EE RQ++W E Q++LR+ LE++ RE+++R Q E+ +
Sbjct: 225 RRRERQHERA-LQEEELQRLQRRWRE-EPREQQQLRR-ELEET-REREQLEQERRE 280
Query: 148 ESPPREPEQLGEDVERRIFTPTSRWRDLEKALSLVPAPSRTO SAHOSRRPHLPMSPTQ 207
+ RRE ++L E ERA ++ + E L R O Q R + +
Sbjct: 281 QQLRRE-ORL-DEEERREQLRRELEIIRERQRLQERREQRLQERREQLKRELE 338
Query: 208 QPALKGRPNSSVEFTYRPRTRRVPKPKRSASFPVTGTSIRLLTMPSLQI3SPANIKK-K 266
+ +QR +E R R + + + +A G S+ R W SA ++ K
Sbjct: 339 ETREREQR---LEQEER-REQLAEVREQAR--ERGESLTR-RWQRLSEAGARQSK 390
Query: 267 VYHMDAQRNQLQLSESELRLPHYLRKALELTTTMM-----ELGALRLQYLCHKY 320
VY +R+ Q L ++ E R R + LE E R Q L +
Sbjct: 391 VYS---RPRRQEQSLRQDQERR-ORQERELEEDQARROQQWQAEESERRRQLSARP 446
Query: 321 IFYRRLQSLRQEAINHVOIMKETEASYKAQNLVI-FLENIQLQSL-RLQAMTDKQKGLE 378
R Q +E Q +E E + + + FLE ++LQ R Q ++ E
Sbjct: 447 SLRER-QLRAGEAQEQEORFREERERREQLQFLKEEQQLRRRAQQLQEDSFOE 505
Query: 379 EKHR 382
++ R
Sbjct: 506 DRER 509

Score = 119 (17.9 bits), Expect = 2.2e-03, P = 2.2e-03
Identities = 79/357 (22%), Positives = 150/357 (42%)

Query: 33 KKWERPVAESLGHKKKKDQEDYFQGGGLQIKFHCSKQLSLESSRQVTSESQEEPWEEETGR 92
++ E+ + + E ++E Q+ + + +G R+ + + EE+ + +
Sbjct: 990 RREQLQERDRKFRREEQLQGE---REERLRQERDRKFRREERQLRQLEEQFRQ 1046
Query: 93 ENRRQLWLEEEEMWQOQKRNWALLEQEHQKLRQWLEDLAREQORRWVOLEKQESPRR 152
E R+ LEE+ + Q+++K L QE K R+ E+ R +Q R QL +E+ + R
Sbjct: 1047 ERDRKFRLEEQ-IRQEAQK-QLRQERDRKFR---EEQRRQERQQLRERDRKFR 1101
Query: 153 EPEQLGEDVERRIFTPTSRWRDLEKALSLVPAPSRTO SAHOSR--RPHLPMSPTQOPA 210

Sbjct: 1102 E EOL ++ E R R L + E L + + + R R + + +
EEQLQEEERERLRQERARKLREE--OLLREEQLLQERDRKFREEQLQEEER 1160

Query: 211 LQKQ---RPHSVFTYRPRTRVPTKPKXSFPVTGTSIRALTWPSLQISPAKIKKV 267
L + Q R + E + R + + + + R + Q + + + +

Sbjct: 1161 LARQERERKLREEQLQEEERLRQERARKLREEQLLQEEQLRQERARKLREE 1220

Query: 268 YHDMERQ-----RKLQLLS--ESELRLPHYLRKALELTTTTLGALRLQYL 316
+ E Q R + QLL EE ELR + + E E LR Q

Sbjct: 1221 QLLAQEEQLRQERDRKFREEQLLREEQLLRERDRKFREEQLLQEEERLRQER 1280

Query: 317 CHYIFYRKLQSLRQEAINHVOIKHTEASYKAOPLYFLEMDRLQ--SLAQANTOKK 375
R + E L + E + + + + E + E RE EQL E + + R R L + E

Sbjct: 1281 ARK--LREEEQLLFEQEEQLRQERDRYRAEQFAREKSRRLERLAQEEQRRR 1338

Query: 376 GLEEKHRE 383
E K RE

Sbjct: 1339 ERERKFRE 1346

Score = 109 (16.4 bits), Expect = 1.9e-01, P = 1.7e-01
Identities = 37/113 (32%), Positives = 60/113 (53%)

Query: 67 KQLESRRQVTSESO--EEPWEETGEMRRLQWLEEEEMMOOROKKWLQEQEK 124
+QL E R + E Q + E EE R + R + EEE+ Q + R + + L QE + KL

Sbjct: 764 QQLAREDRKFREEQLLQEEERLRQERARKLREEQLLQEEER--RLRQERERKL 822

Query: 125 ROWNLEDLAREQORRMVQLEKEQSPREPEOLGEDVERRI FTPTSRWRDLKAE 179
R + E L + E + + + + E + E RE EQL E + + R R L + E

Sbjct: 823 REE--EQLQEEERLR--RQERERKLREEQLLQEEERLRQERERKLREE 872

Score = 107 (16.1 bits), Expect = 3.0e-01, P = 2.6e-01
Identities = 35/109 (32%), Positives = 61/109 (55%)

Query: 71 LESSRQVTSQEPWE--EFGREMRQL---WLEEEEMMOOROKKWLQEQEK 126
L Q + ES + EE + E + + + R + + EEE+ Q + R + + L QE + KL +

Sbjct: 742 LREEQLLQEEERLRQERERKLREEQLLQEEERLRQERERKLREE--RLRQERERKL 800

Query: 127 WNLLEDLAREQORRMVQLEKEQSPREPEOLGEDVERRI FTPTSRWRDLKAE 179
E L + E + + + + E + E RE EQL + + E R R L + E

Sbjct: 801 E--EQLQEEERLR--RQERERKLREEQLLQEEERLRQERERKLREE 850

Score = 104 (15.6 bits), Expect = 9.4e-02, P = 9.0e-02
Identities = 84/339 (24%), Positives = 149/339 (43%)

Query: 67 KQLESRRQVTSQEPWEETGEMRRLQWLEEEEMMOOROKKWLQEQEK 123
+QL E + + EE EE RE R + L + LEEKE Q + R + + L E + + + +

Sbjct: 451 RQLRAEERQEQERFREE---EQRRERRLQELQFLQEEERLRQERERKLREE 507

Query: 124 LACWNLEDLAREQORRMVQLEKEQSPREPEOLGEDVERRI FTPTSRWRDL 175
R + + + Q RW QL + E + R + P EQL E + E R + R R +

Sbjct: 508 ERNRQEQERPGQTRW--QLQEEARARHTLYKPGQEQLEEEELQERERERER 566

Query: 176 EKAEKSLVAPSRTOASQSRAPLWSPSTQOPALGKORPMSSVFTYRPR---RRV 231
+ E L + + + R + + Q + L + R + E + R RR

Sbjct: 567 REE--KLQEEDEKRRERERQYRELEELQEEQL--RDRKLREEQLLQEEERLRQ 624

Query: 232 PTKFK---KSASFPTGTSIRALTWPSLQISPAKIKKVYHDMERQ---NLQLLEE 285
+ K + + + R + L + + + + + E + R K QLL E

Sbjct: 625 ERERKLREEQLLQEEERLRQERERKLREEQLLQEEERLRQERERKLREEQLLQER 684

Query: 286 SELALPHYLRKALE-----LTTTTLGALRLQYLCHYIFYRKL--SLAQEAINH-- 337
E RL R + L L EL R + L + RR Q LRQ +

Sbjct: 685 EERLRQERARKLREEQLLQEEERLRQERERKLREEQLLQEEERLRQERDRKLRE 744

Query: 338 --QIMKTEASYKAOPLYFLEMDRLQSLAQANTOKKGLEEKHRECL 385
Q + + E + E + + E + L + R + + + + L + E + E L

Sbjct: 745 EEQLQEEERLRQ---EREQLRERDRKFREEQLLQEEERLR 789

Score = 103 (15.5 bits), Expect = 1.2e-01, P = 1.1e-01
Identities = 42/152 (27%), Positives = 74/152 (48%)

Query: 36 ERPAESLGHKDKQEDYFKQGLQIFHCQSQLESRRQVTSQEPWEETGEMR 94
ER + K + + + E + + + + + L E + + E QE E + RE

Sbjct: 835 ERLRQERERKLREEQLLQEEERLRQERARKL--REEQLLQEEERLRQERDRKLRE 893

Query: 95 RRLWLEEEEMMOOROKKWA-----LQEEHOKLROWNLEDLAREQ---RRVQ--LEKE 146
+ L EE + E + Q + R + K LL + + E + LR + E RE + + RR Q L + E

Sbjct: 894 EQLLQEEERLRQERDRKLRREEQLLQEEERLRQERERKLREEQLLQEEERLRQER 953

Query: 147 QESPRREPEOLGEDVERRI FTPTSRWRDLKAE 179
+ RE EQL + + E R R L + E

Sbjct: 954 RARKLREEQQLQEREZERLRARQERARKLREE 986
Score = 103 (15.5 bits), Expect = 7.8e-01, P = 5.4e-01
Identities = 31/91 (34%), Positives = 52/91 (57%)
Query: 67 KQLSLESSRQVTSQEEFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQ 126
++L E R++ E Q EE+ R+ R + EEE++ Q-R+++ L QE KLR+
Sbjct: 642 QELRQERERKLREEEQLLRREEQLRQERERKLREEEQLLQEREZE-RLRQERARKLAE 700
Query: 127 WNLEDLAREQORRWVLEKEQESPRPEQL 157
E L R+++ +L +E+E RE EQL
Sbjct: 701 E--EQLLQEEQ---ELRQERERKLREEQL 726
Score = 101 (15.2 bits), Expect = 2.0e-01, P = 1.8e-01
Identities = 38/111 (34%), Positives = 57/111 (51%)
Query: 72 ESSRQVTSQEEFPWEE-EGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQWNLE 130
E R++ E Q EE E RE R+L EEE++ Q-R+++ L QE KLR+ +
Sbjct: 931 ERERKLREEQLLRREEQLRAREARKL-REEEQLLQEREZE-RLRQERARKLREEE-Q 987
Query: 131 DLAREQORRWVLEKEQESPRPEQLGEDVERRITPTSRWDLKAEKSL 182
L RE+Q +L +E++ RE EQL ++ E R R L + E L
Sbjct: 988 LLRREEQ---ELRQERDRKFREEEQLLQEREZERLRQERDRKFREEERQL 1035
Score = 101 (15.2 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 33/108 (30%), Positives = 56/108 (51%)
Query: 72 ESSRQVTSQEEFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQWNLE 131
E R++ E Q EE E R+ R + EEE++ Q-R+++ L QE KLR+ E
Sbjct: 841 ERERKLREEQLLRREEQLRQERARKLREEEQLLQEEQ---LRQERDRKLREE--EQ 895
Query: 132 LAREQORRWVLEKEQESPRPEQLGEDVERRITPTSRWDLKAE 179
L R+++ +L +E++ RE EQL ++ E R R L + E
Sbjct: 896 LLRREEQ---ELRQERDRKLREEEQLLQEREZERLRQERERKLREE 940
Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 32/97 (32%), Positives = 50/97 (51%)
Query: 72 ESSRQVTSQEEFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQWNLE 131
E R+ E Q EE E R L EEE Q +++ L QE + KLR+ E
Sbjct: 578 EKRRQERERQYRELEELRQEEQLRQERDRKLREEEQLLQEREZERLRQERERKLREE--EQ 635
Query: 132 LAREQ-----ORRWVLEKEQESPRPEQLGEDVERRI 165
L R++ Q R +L +E++ ARE +L ++ ER++
Sbjct: 636 LLRQEEQLRQERERKLREEEQLLRREEQLRQERERKL 674
Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 34/111 (30%), Positives = 58/111 (52%)
Query: 67 KQLSLESSRQVTSQ--EFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKL 124
++L E R++ E Q +E EE R+ R + EEE++ +Q +++ L QE + KL
Sbjct: 664 QELRQERERKLREEEQLLQEREZERLRQERARKLREEEQLLQEEQ---LRQERERKL 720
Query: 125 RQWNLEDLAREQORRWVLEKEQESPRPEQLGEDVERRITPTSRWDLK 177
R+ + L RE+Q L +E++ RE EQL ++ E R + L +
Sbjct: 721 REEE-QLLRREEQL---LRQERDRKLREEEQLLQEREZERLRQEREQQLR 768
Score = 98 (14.7 bits), Expect = 2.6e+00, P = 9.2e-01
Identities = 37/146 (25%), Positives = 77/146 (52%)
Query: 20 EPVLLPVDRFPKMERFVAESLGHKKKQDEDFQGGQJQKPHCSKQLSLESSRQVTS 79
E LL ++ ++ ER + E +E+ ++ K +QL + +++
Sbjct: 655 EQLLRREEQLRQERERKLREEEQLLQEREZERLRQERARKLREEEQLLRQEEQLRQ 714
Query: 80 ESQEEFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQWNLE-LAREQOR 138
E + + EEE + +R+ L +E ++ +++ LL++ +S-LR+ E L RE+ R
Sbjct: 715 ERERKLREE--QLLAREEQLLRQERDRKLREEEQLLQEREZERLRQEREQQLRERDR 772
Query: 139 RWVLEKEQESPRPEQLG-EDVERRI 165
++ E+EQ RE E+L ++ ER++
Sbjct: 773 KP--REEEQLLQEREZERLRQERERKL 798
Score = 97 (14.6 bits), Expect = 3.3e+00, P = 9.6e-01
Identities = 38/129 (29%), Positives = 63/129 (48%)
Query: 72 ESSRQVTSQ--EFPWEEFGREMRQLWLEEEEMWQORQKWKALLEQEHQEKLRQWNLE 129
E R++ E Q +E EE R+ R + EEE++ +Q +++ L QE KLR+
Sbjct: 817 ERERKLREEEQLLQEREZERLRQERERKLREEEQLLQEEQ---LRQERARKLREE-- 871
Query: 130 DLAREQORRWVLEKEQESPRPEQLGEDVERRITPTSRWDLKAEKSLVPAPSRT 189
E L R+++ +L +E++ RE EQL E+ + R R L + E L+

Sbjct: 872 EQLLRQEEQ--ELRQERDRKLREEEQLLAQEEQEL--RQERDRKLREE-QLLQSEEE 925

Query: 190 QSAHQSRPHL 200
+ Q R L

Sbjct: 926 RLRRQERERKL 936

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 41/132 (31%), Positives = 69/132 (52%)

Query: 46 KKKDDQEDYFQGGGLQI-KFHCSKQLSLESSQVTSSESQEEPWEEFGREMRRLWLEEE 104
+++ QE F + Q+ + ++OL E S Q E + E+ G+ R QL +EE

Sbjct: 473 RERRQELQFLEEEQLQRRERAOQLQESDSQEDRERARRQQRPGQYWRWQL--QEE 529

Query: 105 MWQORQKQWALLEQEHQEKLRQWNLQDLAREQQRWVLEKEQESPREPEQLGEDVERR 164
++R +A Q QE+LR+ E+L RE++R+ E+E+E E O ED +RR

Sbjct: 530 AQRRRHTLYAKPGQ--QEQLREE--ELQREKRRQ----ERERERREEELQREDEKRR 581

Query: 165 IFTPTSRWRDLEK 177
++R+LE+

Sbjct: 582 RQERERQYRELEE 594

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 35/138 (25%), Positives = 76/138 (55%)

Query: 28 DRAPFKKWERPVAESL-GHKDDQEDYFQGGGLQIKFHCSKQLSLESSQVTSSESQEPW 86
+R++ + E E L R +++E Q+ + ++ L Q+ + ++E

Sbjct: 586 ERQYRELEELRQEEQLRDRKLREEEQQLQREERELRQERERKLREEEQLLRQEEQE-L 644

Query: 87 EEEFGREMRRLWL--EEEMWQORQKQWALLEQEHQEKLRQWNLQDLAREQQRWV 143
+E R++R + L EE+E+ Q+R++K L +E Q L++ E L R+++ R +L

Sbjct: 645 RQERERKLREEEQLLRQEEQLRQERERK---LRKEEO-LQREERELRRQERAR--KL 698

Query: 144 EKQESPREPEQLGEDVERR 165
+E++ R+E ++L ++ ER++

Sbjct: 699 REEEQLRQEEQLRQERERKL 720

Score = 95 (14.3 bits), Expect = 5.2e+00, P = 9.9e-01
Identities = 59/282 (20%), Positives = 121/282 (42%)

Query: 20 EPULLPLVDRFPFKKWERPVAESLGHKDDQEDYFQGGGLQIKFHCSKQLSLESSQVTS 79
E LL ++ ++ ER + E + +E+ ++ K +OL + +++

Sbjct: 655 EQQLRREEQELRQERERKLREEEQQLQREERELRQERARKLREEEQLLRQEEQELRQ 714

Query: 80 ESQEEPWEEFGREMRRLWLEEEEMWQORQKQWALLEQEHQEKLRQWNLQDLAREQQR 138
E + + EEE + +RR+ L +E ++ ++ LL++ +E+LR+ E L RE+ R

Sbjct: 715 ERERKLREEE--QLLRREEQLRQERDRKLREEEQLLQSEEEERLRQEREQQLRREDR 772

Query: 139 RWVLEKEQESPREPEQLG-EQVERRIFTPTSRWRDLEKAEKLSLVPAPSRQSAHQ--S 195
++ E+EQ RE E+L ++ ER++ ++ E+ L + + Q

Sbjct: 773 KF--REEEQQLQREERELRQERERKLREEEQQLQREERELRQERERKLREEEQLLQ 830

Query: 196 RRPHLMSPTQOPALGRQPMSSVEFTYRPTRVPTPKKSASFVVTGTSIRALTWPS 255
R + + ++ L ++ + E R R ++ + +

Sbjct: 831 EREERLRQERERKLREEEQLLRQE-EQLRQERARKLREEEQLLRQEEQLRQERDRK 889

Query: 256 LQISPAKIKKKVYHNDLEAQRK---NLQLSSEELRPLHYLSKAL 299
L+ +++++ + E RK QLL E E RL + + L

Sbjct: 890 LRREEQLRQEEQLRQERDRKLREEEQLLQSEEEERLRQERERKL 936

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 35/116 (30%), Positives = 59/116 (50%)

Query: 72 ESSQVTSSESQEEPWEEFGREMRRLWLEEEEMWQORQKQWALLEQEHQEK-----L 124
E +R++ E Q EE+ R+ R + + EEE++ Q+R+++ L QE R L

Sbjct: 977 ERARKLREEEQLLRREEQELRQERDRKFRFEEQQLQREEE-RLRRQERDRFREEERQL 1035

Query: 125 RQWNLQDLAREQQRWVLEKEQESPREPEQLGEDVERRIFTPTSRWRDLEKAEKLSL 182
R+ LE+ R+++ R +LE EQ +E +OL R F + R ++ E L

Sbjct: 1036 RQEELEEFQQRDRRFLRLE-EQIRQEREEQLRQERDRKFRFEEQQRARRQERQOL 1092

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 51/166 (30%), Positives = 76/166 (45%)

Query: 67 KQLSLESSQVTSSESQ--EEPWEEFGREMR-RQWLEEEEMWQORQKQWALLEQEHQEK 123
++L E R+ E Q +E EE R+ R R+L EEE++ + Q+ L QE+

Sbjct: 1250 QELRRERDRKFRFEEQQLQREERELRQERARKLREEEQLLFEEQEEQRL----RQER 1305

Query: 124 LRQWNLQDLAREQQRWVLEKEQESPREPEQLGEDVERRIFTPTSRWRDLEKAEKLSL 182
R++ E+ ARE++ R +LE+E R+E EQ R F R E+ E

Sbjct: 1306 DRYRAREQFAREKSR--RLREL---RQEEQRARRERARKFRFEEQLRQEE-EQRR 1359

Query: 183 VPAPSRQSAHQSRPHLPMSPSTOOPALGKORPMSSVEFTYRPRTRVP 232
R QRR L P T+Q A R E+ R++ P
Sbjct: 1360 AQLRERQFREQSRRQVL--EPGTRQFARVPVRSSPLYEQRSQYRP 1407
Score = 93 (14.0 bits), Expect = 8.3e+00, P = 1.0e+00
Identities = 41/145 (28%), Positives = 72/145 (49%)
Query: 28 DRFFFKWERPVASLGHKKQDQDYFQKGLQTKFNCSKQLSLESSQVTSQEEPW- 86
+RR ++ ER + E + + Q + + Q + L R + QE+ +
Sbjct: 408 ERRQQRERERELEQARRQQQWQAEESERRRQ-RLSARPSLRERQLRAEERQEQQRFR 466
Query: 87 -EEETGEMRRQL-WLEEDMWOQKQKQWALQEQ--HQEKLRQWNLQDLAREQQRWVQ 142
EEZ RE R++L +LEEE Q+R++ L E++ ++ R+ ++ Q RM Q
Sbjct: 467 EEEQRERRRQLQGLEEEQQLRRRAQQLQEDSFOEDREARRAQEQRPQOTWRW-Q 525
Query: 143 LEKEQSPRR----EP---EQLGEDVE 162
L++E + R +P EQL E+ E
Sbjct: 526 LQEEAQRNHTLYAKPGQEQLEEE 552
Score = 91 (13.7 bits), Expect = 2.4e+00, P = 9.1e-01
Identities = 38/110 (34%), Positives = 57/110 (51%)
Query: 72 ESSRQVTSQEEPWEE-EPGEMRRQLWLEEDMWOQKQKQWALQEQHQEKLRQWNL- 129
E R++ E Q EE E RE R+L EEE+ Q+R++ L QE KLR+
Sbjct: 931 ERERKLEEQQLRRQEQELRRARKL-REESQLQEREE-RLARQERARKLREEEQ 988
Query: 130 -----EDLAREQQRWVQLEKEQSPREPEQLGEDVERRI FTPTSRRDLEKAE 180
+L +E+ R++ E+EQ RE E+L R F R L + EL
Sbjct: 989 LRAREQELRQERDKF--REEDQLQEREERLQERQERDKFREER--QLRQEL 1040
Score = 89 (13.4 bits), Expect = 2.2e+00, P = 8.9e-01
Identities = 35/138 (25%), Positives = 65/138 (47%)
Query: 82 QEEPWEEETGEMRRQLWLEEDM--WQORQKQWALQEQHQEKLRQWNLQDLAREQQR 139
Q E++ E+R + + +E E WQ+ + + L E+ E Q K R + + +R+ + +
Sbjct: 111 QNRQEQQRRTFLRDQFQEPERRRQKQEQERELAEKEEQKRRERFQNTSROYRDK 170
Query: 140 WVQLEKEQ-ESPREPEQL----GEDVERRI FTPTSRRDLEKAE LVPAPSRQSAHQ 194
+L+ + + E R E EQL G D E F + R E+ EL Q +
Sbjct: 171 EQRLQEQLEERAEELQARRKGRDAEE--FIEEQARRRQEQELKR-ELREEEQORRE 227
Query: 195 SRRPHLPMSPSTOOPALGQR 215
R H ++ L +R
Sbjct: 228 RREQERALQEEELQRR 248
Score = 50 (7.5 bits), Expect = 2.2e+00, P = 8.9e-01
Identities = 34/160 (21%), Positives = 67/160 (41%)
Query: 325 RLQSLQEAINHVOINKETEASYKAQNYIFLENIDRL-QSLRLQAWTDKQGLEEKHRE 383
R + R+E Q+ +E E + + LE +R Q LR + + + + E+ R
Sbjct: 245 QRWRKEEPREQQLARELEKIRERQR--LEQERREQLRREORLEQERREQLAR 301
Query: 384 CLSSMVTMPKQLQENRVHLNIP-EVTSPPKPKCKLPASPRHIRPSGPTYKPLSRH 442
L + +L E + E + K+L R R + + L+
Sbjct: 302 ELEIRERERQLEQERREORLEQERREQLKRELEIRERERQLEQERREOLLAEEV 361
Query: 443 ACVPLQARQOQKQHEAVNKEVASSYAEKXTPASLPDQ 484
+ AR++G+ + W+ ++ S + A + E S PR Q
Sbjct: 362 R----EQAREGESLTRWQROLESEAGARQSKV-YSRPRQ 398
Score = 40 (6.0 bits), Expect = 1.9e-01, P = 1.7e-01
Identities = 32/115 (27%), Positives = 47/115 (40%)
Query: 276 RKNLQLLSESELRLPHYLRKAL--ELTTTTELGAIRLQYLCHRYIFYRL-OSLRQE 332
R+ QLL E E RL R++L E E E LR Q R+ +L Q +E
Sbjct: 959 REEQQLQERERELRQERARKLREEQLKREKQLR-QERDRFREEDQLQEREE 1017
Query: 333 AINHVI---NKETEASYKAQNYIFLENIDRLQSLRLQAWTDKQ-KGLEEKHRE 383
+ + +E E + Q L F + DR L Q +K+ K L + R+
Sbjct: 1018 RLARQERDRKFRERQLRQLEEQFRQERDRKFRLEEQIROEKEERQLRQERD 1073
Score = 37 (5.6 bits), Expect = 1.6e+00, P = 7.9e-01
Identities = 27/108 (25%), Positives = 43/108 (39%)
Query: 276 RKNLQLLSESELRLPHYLRKAL--ELTTTTELGAIRLQYLCHRYIFYRLQSLRQE 332
R+ QLL E E RL R+ L E E LR Q K R + L QE
Sbjct: 775 REEQQLQERERELRQERARKLREEQLQEREEELRQERERKL---REEQLQEQ 831
Query: 333 AINHVOINKETEASYKAQNYIFLENIDRLQSLRLQAWTDKQ-KGLEEKHRE 383
+E E + + + E L+ R + + + + L + +E
Sbjct: 832 REERLQARQERERKLREEQLRQEE-QLRQERARKLREEQLRQEQEQE 881

Report for DKFZphtes3_50n23.1

(No Prosite data available for DKFZphtes3_50n23.1)
(No Pfam data available for DKFZphtes3_50n23.1)

WO 01/12659

PCT/IB00/01496

DKFZphes3 6b21

group: testes derived

DKFZphes3 6b21 encodes a novel 781 amino acid protein without similarity to human KIAA0256 gene product.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0256

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="J56.3 cR from top of Chr9 linkage group"

Insert length: 3360 bp

Poly A stretch at pos. 3314, polyadenylation signal at pos. 3300

```
1  GGCAAGCCGA CGGCCCGCTG CTGGCCCTCG TGACGGGGCC TCCTCCGCGC
51  CTCGCGGCAT GCGCTCGGAG GGGCCCGCGG AGCCGGAAG CGAGGGCATC
101  AAGTATATCAG CAGATGTCTA ACCATTGTGC CCGCAGATTG CCGGCTGCAA
151  TGTGGCATGG TTAGAGTCTT CAGAAAGCAT GTCTCTCCCC AGCTCTGCAG
201  CCACATACTA TCGTTTGTGT CAGGAAGCAC CAGTGACAGA AATGTTTACT
251  CAGTGGCTGG CTGCGAGTAT CTTTATAACC AACCCAGTTG TTACCGAGGT
301  TTTCAACACG TGAGCATCTG AATGAGAGAC ACATGGCCCTG TCCACACAGA
351  AATGAAGACT CTGTTTAAGA AGAAACCTTA TGATGAGAAA AAAACGTATG
401  ATCAGCAAAA GTTTGACAGT GAAAGGGCTG ATGGAACAT ATCATCTGAG
451  ATAAATACAG CTAGAGGTTT ACATCAATTT TCCATTAGAG CTGAGAATAG
501  TTTGAATATCA GATGTTTACC ATAGCGCAAC AGACAGGAAA TCAGAATCA
551  TTGCAAAAAA TGTATCTACC TCCAACCTGT AGTTTGAATT TACCACACTG
601  GACTTTCCTG AACTGCAAGG TGCAGAGAAC AATATGTCAG AGATACAGAA
651  GCAACCCCAAG TGGGGACCTG TCCACTCTGT CTCTACCGAC ATTTCCTCTC
701  TAAGGAGAGT AGTAAACCA CCGCGAGTGT TATCAAGGGG TGAATATGTG
751  GTGAATAATA ACCCAATGCA ATCTGTAATC GCTAATGCCG CTACCAATTG
801  TCCTTCAATG ACAAGAGAGT TATCTTGGAC ACCAATGGGT TATGTTGTTT
851  GACAGACATT ATCTACAGAA CTGTCAGCAG CCGCTAAAAA TGTACTCTCT
901  ATGATAAATC TAAGAGCATC TCGTCTCATC GCAGATCTTA AAAATGTTAG
951  TATACCATCT TCTGAAGCTT TATCTTCGGA TCCTTCTTAC AACAAGAAAA
1001  AACACATTAT TCATCCTACC CAAAAGTCTA AAGCATCACA AGTAGTGAC
1051  CTTGAGCAAAA ATGAAGCTTC AAGAAAGAA ATAGAAAAAG AAGAAAAATC
1101  TACATCAAAA TATGAAGTCC TGACAGTTCT AGAGCTCTCA AGATTGAAG
1151  ATGCCGAGGA ATTTCCCAAC CTGGCAGTTG CATCTGAAG AAGAGACAGA
1201  ATAGAGACAC CGAAATTTCA ATCTAAGCAG CAGCCACAGG ATAATTTTAA
1251  AAATAATGTA AGAAGAGGCC AGCTTCCAGT GCAGTTGGAC TTGGGGGGCA
1301  TCCTGACACG CCGTGGAAGG AAGCAGCACT CTCAGATCTC AAGACGTCC
1351  TCCAACCCAG TGGTAGTCTC AGTTGGAGCA GTGCCAGTCC TTCCCAAGA
1401  ATGTGCATCA GGGGAGAGAG GCGCCCGCAT GAGTCAATG AAGACCCCGC
1451  ACAATCCCTT GGACTCCAGC GCGCCCACTG TGAAGAAGG GAAGCAGAGG
1501  GAGATCCCCA AGCCCAAGA GCCCAACTCA CTGAAGAGA TTATTTGAAA
1551  AGAACGCCAA GAGAGAAAGC AGCGTCTCCA AGAAATGCT GTAGTCCAG
1601  CTTTACCAGG TGATGACACA CAAGATGGAG AGAGTGGTG TGATGACCAG
1651  TTTCCCGAGC AGGCAGAGCT GTCAGGGCCA GAGGGGATGG ACGACTGAT
1701  CTCCACTCTT TCGGTTGAGG ACAGCTCTCA AGAGCCACCA GGCACAGGC
1751  TCCAGAGGGA CACAGAGGCC TCCGACCTTG TCTCCAATCA CACCACCTTC
1801  CCTAAGATCC ACAGCCGAGC ATTCAGGGAT TACTGCAGCC AGATGCTTAG
1851  TAAAGAAAGT GATGCTTGTG TTACCGACCT ACTCAAGAA CTGTCGCTTT
1901  TCCAAGACCG TATGTACAGG AAAGATCCAG TCAAGGCCAA GACTAAAGCT
1951  CGACTTGTGT TGGGGTGAAG GGAAGCTCTG AAACACTGTA AGCTCAAAAA
2001  ACTGAAATGT GTCAATTATT CTCCCAACTG TGAGAAGATA CAGTCAAAAG
2051  GTGGGCTGGA TGACACTTTG CACACAATTA TTGATTATGC CTGTGAGCAG
2101  AACATTCCCT TTGTGTTTGG TCTAACCCGG AAAGCTCTGG GCGCCAGTTT
2151  GAATTAAGCA GTTCTCTCTA GTGTGCTGGG GATCTTCCGC TATGATGGGG
2201  CCCAGGATCA GTTCCACAAG ATGCTTTGAG TGACAGTGGC GCGCCGACAG
2251  CGGTACAAGA CCATGCTGGA GAATGTGCAG CAGGAGCTGG TGGGAGAGCC
2301  CAGGCGCTCAG GCACCTCCCA CCTAGCCAC ACAGGGCCCC AGCTGCCCTG
2351  CAGAGATATG CCCCCAGGC CTGAAGAAAG AAGAGAGGCC ACCTACAT
2401  GAAATCTGGA AAAACATCT GGAAGCATAC AGTGGATGTA CCTGGAGCT
2451  AGAAGAAATC TTGAGGCTT CAACCTCTCA AATGATGAAT TTGAATTTAT
2501  GAGAGTTCTT GCTGTGTGTG CTGTATTTTG GGTAAAGAGG GGAGGTCTGA
2551  AAAAGACTTT GGGGCTTTT CTCTGTTTGT TCATGCAAT GTAAATTTGT
2601  TAACGTCTGA ATCTGGAAAT TGATCAGCAT TAAAGGGCAC ATGAGCAGT
2651  GTCTCGAGGC GTTCAGTGCT GCGGAGCCTG TTAAGGTGTA CTCAGATGTG
```

WO 01/12659

PCT/IB00/01496

```
2701 CAGGCTGTAA TCTTCTTAA AAGCCTGGT ATACAGCTCT GGCTTTCTGA
2751 GCACACTAGG GATCTGGAAA ATACTGGAAA ATGTGACTCT TAGAATACTT
2801 TGGCTGCTAA GGAAGCTTCC TCTCCATTGC AGAATAGCTG AGCCAAGTGA
2851 GTGAGTTTTC AGAAGCAGG TGGTGAGCTC CTGCTGCTG GAGGTTTCCA
2901 TGGAGGGCCA TTCTGCCCC GCAAGAGCAC CTCTGTGAG GAGGCTACTT
2951 GGCAGAAGGG TGCAGGGCTG CTGGTGTGAG AGCAAGAGGG CTACAGGGAA
3001 AGGGCCCTTT CTCAGGGGAT GTAGCTTTT TAAAGATT TGGACACTT
3051 GGAGGATTTC CTAAATGAG CCTCAGAGG AAAATTGUTT TTCTAACCTG
3101 TGACTTTTTC AAATGAATTA TTCTTTTTC TCTTTATTT TCAAGAAC
3151 AATGTGTATT GAAGTACCTA GATTGTGTTG ATAATCAACA AATCTTCTCT
3201 TTTTCAATGA ACATATTCTG AATGTGGTTT CTGCTTTAGA CCAGGAGGAC
3251 AGAGTTTGGT TTCATATTTT CCTGTAAAT AACAGGGCTT ATTATTTTTA
3301 AATAAAGAGT AATTATTAAA AAAAAA AAAA AAAA
3351 AAAAAA
```

BLAST Results

Entry NS773347 from database EMBL:
human STS W1-18160.
Score = 813, $E = 2.9e-30$, identities = 167/171

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 157 bp to 2499 bp; peptide length: 781
Category: similarity to known protein

```
1 MVRVLRSMCL POLCSHILSV CSCTTSDRNV YSVPGSOVLY NQPSYRGFO
51 TVKRRMENTC PLPOEMHALF KRTTYDEKXT VQOKTOSER ADGTSSSEIK
101 SARGSHHLSI YAENSLKSDG YHKRTDRKSR IIAKNVSTSK PEFEITTLDF
151 PELQGAENNM SEIQKQPKWG PVHVSSTDIS LLREVVKPAA VLSKEIVVK
201 NNPNESVTAN AATNSPSCYR ELSWTFMGYV VQRTLSTELS AAPNVTSMI
251 NLKTIASSAD PNVVSIPEE ALSOSPVMK KWHIINPQR SKASQGSOLE
301 ONEASRKNKK KKEKSTSYE VLTQEPPI EDAEEFNL VASERRDRIE
351 TPRFOSKQOP QDNFNHVKK SQLPVOLDLG GMLTALEKQ HSQAKQSSK
401 PVVSVGAVP VLSKECASOE RGRHMSQMT PINPLOSSAP LHKRQOREI
451 PKAKKFTSLK KILKEDGER KQLQENAVS PFTSDDTQD GESGGDOOP
501 EQNELSGPEG NDELSTPSV EDKSEEPPT ELORDTEASH LAPNHTTFPK
551 IHSRRFRDYC SQMLSKEVDA CVTDLKELV RFQORHYQKD PVKATKRRAL
601 VLGLREVLRH LKLKELKCVI ISPNCEKIQS EGGLODTLHT IIDYACEQNI
651 PFVFLNKRKA LORSLEKAVP VSVVGIPTD GAGQOYHNV ELTVAAQRAY
701 KTNLENVOOE LVGEPRPQAP PSLPTGCPSC PAEDGPPALK EKEEPHYIEI
751 WKHLEAYSG CTLELESLE ASTSQNNLN L
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for ORFzptes2_6b21, frame 1

SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256., N = 1, Score = 786, $E = 3.6e-78$

TREMBL:PFMAL3P3 15 gene: "MAL3P3.15"; Plasmodium falciparum MAL3P3, N = 2, Score = 161, $E = 5.1e-10$

TREMBL:RNNFLH.1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 150, $E = 9.1e-07$

>SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256.
Length = 635

HSPs:

Score = 786 (117.9 bits), Expect = 3.6e-78, $E = 3.6e-78$
Identities = 190/424 (44%), Positives = 263/424 (62%)

Query: 369 KESQLPVQLDGLGMLTALEKQHSQHAQ---SSKPVVSVGAVPVLSKECASGERGRMS 426
KK++ PVQLDGL ML ALEK+Q + A+Q +++P+ +V + + + S
Sbjct: 16 KKNKTPVQLDGLGMLAALKQQAQKARQITNTRPLSYTVVTAASFHTKOSTNRKPLTKS 75
Query: 427 Q-MKTPHNPLDSSAPLMKKQKQREIPKAKKPTSLKKIILKERQERKQRLQENAVSPFTS 485
Q T N +D ++ KKGK++EI K K+PT+LKK+ILKER+E+K RL + S
Sbjct: 76 QPCLTSFNSVDIASSKARKGREKEIAKLKRPALKKVKILKEREKGRITVD--HNLGGS 133
Query: 486 DOTQDGESGGDQDFEQAELSGPEQMDLISTPSVEDKSEPPG--TELQDTEASML-- 541
++ + D P++ G+ + S S+ S+ P T + + + AS
Sbjct: 134 EEPTEHMLDPIDDLPOEIVSOEDTGLS-MPSDTLSPASQNSPYCHTVPVSGSPASSGIG 192
Query: 542 APN-HTTFPKIHSRRFRDYCSQMLSKVDACVTLKELVRFQDRMYQKDPVKAKTKARL 600
+P +T KHS+RFR+YC+Q+L KE+D CVT LL+ELV FQ+R+YQKDPV+AK +RRL
Sbjct: 193 SPMASSSTITIKSKRFREYCNQVLCKEIDECVTLQLQELVSFQERIYQKDPVRAKARRL 252
Query: 601 VLGLREVVLKHLKKLKCVIISPNCEKIOSKGLDQDYLHTIIVACEQNIIPVFFALNRKA 660
V+GLREV KH+KL K+KCVIISPNCEKIOSKGLD+ L+ +I A EQ IIPVFFAL RKA
Sbjct: 253 VNLGREVTKHKLKIKCVIISPNCEKIOSKGLDEALYNVIAHAREQEIIPVFFALGRKA 312
Query: 661 LGRSLNKAVPVSVVGIFSYDGAQDQTHMVELTVAARQAYKTMLEHVQOELVGEPRP--- 717
LGR +NK VPVSVVGIF+Y GA+ F+K+VELT AR+AYK M+ ++OE E
Sbjct: 313 LGRCVNKLVPVSVVGIFNYFGAELFNKLVELTEARAYKDMVAAMEQEQAEALKNVK 372
Query: 718 QAPPSLP-TQGPS-----CPAEDGPPALKEKEEPHYIEMKKHLEAYSGCTL---ELE 766
+ P + ++ PS C P + E E Y W+ +E G E E
Sbjct: 373 KVPNHMGHSRNPAAASAIAPCSVISEP--LSEVNEKEYETNRNMVETSDGLEASENEKE 430
Query: 767 ESLEASTSQ 775
S + STS+
Sbjct: 431 VSCKHSTSE 439

Pedant information for DKF2phtes3_6b21, frame 1

Report for DKF2phtes3_6b21.1

[LENGTH] 781
[MW] 87393.44
[PI] 8.94
[HOMOL] SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256. 4e-75
[PROSITE] MYRISTYL 4
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 16
[PROSITE] TYR_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 16
[PROSITE] ASN_GLYCOSYLATION 6
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 8.45 %

SEQ MVRVLRSLNCLPQLCSHILSVCSGTTSDRNVSYPGSOYLYNQPSYRGFQTVKRNEMTC
SEG
PRD ccc
SEQ PLPQEMKALFKKKTYDEKKTYYDQKFDSERADGTISSEIKSARGSHHLSIYAENSLKSDG
SEG
PRD ccc
SEQ YKERTDKSRILAKNVSTSKPEFETTLDFPELQGAENHSEIQKQPKWGPVHSVSTDIS
SEG
PRD ccc
SEQ LLREVVKPAAVLSKGEIVVKNPNHESVTANAATNSPCTRELSTWTPMGVYVRQTLSTELS
SEG
PRD hhh
SEQ AAFKNVTSMILKTIASSADPNVSI PSSEALSDPSYNKEKHIIHPTQKSKASQGSLE
SEG
PRD ccc
SEQ QNEASRNKKKKEKSTSKYELTVQEPRIEAEFPNLAVASERRDRIDTPKFSQKQPP
SEG
PRD hhh
SEQ QDKFNKNVKSQLPVQLDGLGMLTALEKQHSQHAQSSKPVVSVGAVPVLSKECASGE
SEG
PRD ccc


```
SEQ      RGRNRSQMKTPHNPIDSGAPLMKGGKQREI PKAKPTSLAKI LKEQRQKQRLQGNASV
PRD      chhhhhtccccccccccccccccchhhhhtccccccccchhhhhtccccccccchhhhhtcccccc

SEQ      PAFTSDDTCECCGCGGQDFPEQALSESGMDGLI TSFVDEKSEEPHGLQDQRT EASH
PRD      cccccccccccccccccccccchhhhhtcccccccccccccccccccccccccccccccccccccc

SEQ      LAPWHITFPKISRFRDYSQMLKEVDACTVLDLKELVRFODNRYQKDPVKATKRRL
PRD      cccccccccccccccccccccchhhhhtccccccccchhhhhtccccccccchhhhhtcccccc

SEQ      VLGLREVLKHLKHLKCVTI SPNCKI QSGKGLDQTLHTI DYACEQNI PFVALNRKA
PRD      hhhhhtcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      LGRSLNKAIVPSVSVGIFSYDQAQOQRHMYVLAARQATMLNVEQELVGEFGRPAQ
PRD      cccccccccccccccccccccchhhhhtccccccccchhhhhtccccccccchhhhhtcccccc

SEQ      PSLPTGPGSPADGPGALKEKEEPHYIEMKHLAYSGCTLESLSEASTSOBNHL
PRD      cccccccccccccccccchhhhhtccccccccchhhhhtccccchhhhhtcccccccccccccc

SEQ      L
PRD      C
```

Prosite for OXKZphtes3_6b21.1

P500001	125-139	ASN_GLYCYSTYLATION	PDOC00001
P500001	159-163	ASN_GLYCYSTYLATION	PDOC00001
P500001	204-208	ASN_GLYCYSTYLATION	PDOC00001
P500001	228-249	ASN_GLYCYSTYLATION	PDOC00001
P500001	263-267	ASN_GLYCYSTYLATION	PDOC00001
P500001	544-558	ASN_GLYCYSTYLATION	PDOC00001
P500004	71-75	CAMP_PROSPHO_SITE	PDOC00004
P500004	425-439	CAMP_PROSPHO_SITE	PDOC00004
P500004	454-458	CAMP_PROSPHO_SITE	PDOC00004
P500005	26-29	PKC_PHOSPHO_SITE	PDOC00005
P500005	51-54	PKC_PHOSPHO_SITE	PDOC00005
P500005	88-91	PKC_PHOSPHO_SITE	PDOC00005
P500005	101-104	PKC_PHOSPHO_SITE	PDOC00005
P500005	115-118	PKC_PHOSPHO_SITE	PDOC00005
P500005	125-128	PKC_PHOSPHO_SITE	PDOC00005
P500005	138-141	PKC_PHOSPHO_SITE	PDOC00005
P500005	288-291	PKC_PHOSPHO_SITE	PDOC00005
P500005	305-308	PKC_PHOSPHO_SITE	PDOC00005
P500005	316-319	PKC_PHOSPHO_SITE	PDOC00005
P500005	343-346	PKC_PHOSPHO_SITE	PDOC00005
P500005	351-354	PKC_PHOSPHO_SITE	PDOC00005
P500005	398-401	PKC_PHOSPHO_SITE	PDOC00005
P500005	408-411	PKC_PHOSPHO_SITE	PDOC00005
P500005	553-556	PKC_PHOSPHO_SITE	PDOC00005
P500005	596-599	PKC_PHOSPHO_SITE	PDOC00005
P500006	24-28	CK2_PHOSPHO_SITE	PDOC00006
P500006	78-82	CK2_PHOSPHO_SITE	PDOC00006
P500006	139-143	CK2_PHOSPHO_SITE	PDOC00006
P500006	146-150	CK2_PHOSPHO_SITE	PDOC00006
P500006	193-197	CK2_PHOSPHO_SITE	PDOC00006
P500006	217-221	CK2_PHOSPHO_SITE	PDOC00006
P500006	297-301	CK2_PHOSPHO_SITE	PDOC00006
P500006	317-321	CK2_PHOSPHO_SITE	PDOC00006
P500006	323-327	CK2_PHOSPHO_SITE	PDOC00006
P500006	354-358	CK2_PHOSPHO_SITE	PDOC00006
P500006	484-488	CK2_PHOSPHO_SITE	PDOC00006
P500006	493-497	CK2_PHOSPHO_SITE	PDOC00006
P500006	506-510	CK2_PHOSPHO_SITE	PDOC00006
P500006	517-521	CK2_PHOSPHO_SITE	PDOC00006
P500006	604-644	CK2_PHOSPHO_SITE	PDOC00006
P500006	702-706	CK2_PHOSPHO_SITE	PDOC00006
P500007	581-588	TYR_PHOSPHO_SITE	PDOC00007
P500007	740-748	TYR_PHOSPHO_SITE	PDOC00007
P500007	740-748	TYR_PHOSPHO_SITE	PDOC00007
P500007	73-82	TYR_PHOSPHO_SITE	PDOC00007
P500008	93-99	MYRISTYL	PDOC00008
P500008	181-184	MYRISTYL	PDOC00008
P500008	380-386	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	633->639	MYRISTYL	PDOC00008
PS00009	421->425	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3_6b21.1)

WO 01/12659

PCT/IB00/01496

DKFiphtes3_6c11

group: signal transduction

DKFiphtes3_6c11 encodes a novel 1025 amino acid protein with similarity to A. ambisexualis antheridiol steroid receptor.

The novel protein is a putative steroid receptor. It shares similarity with yeast YNL132w and contains the ATP/GTP-binding site motif A (P-loop) and RGD site, similar to the A. ambisexualis antheridiol steroid receptor.

The new protein can find application in modulating/blocking the expression of genes controlled by this receptor.

strong similarity to YNL132w

strong similarity to S.pombe/YDK9_SCHPO, S.cerevisiae/YNL132w, C.elegans/F55A12.0

Sequenced by BMF2

Locus: unknown

Insert length: 3966 bp

Poly A stretch at pos. 3890, polyadenylation signal at pos. 3873

```

1 GCTGTGCCCT CTCTTCGGA GTTCTCCGT GCTCCACGT GCTTCCCTT
51 CTCCACTGGC TGGATCCCC CGGGCTCGGG GCGCAGTAT AATTTTCAC
101 CATGCATCGG AAAAAGTGG ATACCGAAT CCGATCTCT ATTGAGATG
151 GAGTACGTGA GCGGCAAGA TCTCTTTTG TTGTAGTTG GATCGAGGA
201 AAAGATCAGG TGTAACTAT TCATCACATG TTATCGAAG CAACTGTGAA
251 GGCTCGGGCT TCACTGCTGT GGTGTTTAA GAAAGAGCTG GGGTTTANGA
301 GTCACTCGGA GAAAGATAT GGCACAGCT AGAAGAAAT AAGAAATGGA
351 ACACCTAGCA TAAAGCAGCA GACGCCCTT GAACCTTCA TAGCAGGCAC
401 AAACATTCCG TACTGCTACT ACAACGAGAC CCACAAAGAT CTGGGCAATA
451 CCTTCGGCAT GTGTGTGCTG CAGGATTTTG AAGCCTTAAC TCCAAACTTG
501 CTGGCCAGGA CTGTAGAAAG ACTGCAAGGT GGTGGGCTAG TGTGATCTCT
551 CTTAGCGGAC ATGAATCTAC TCAAGCAATT GTACACAGTG ACTATGGATG
601 TGCATTCCAG GTACAGAAGT GAGGCCCATC AGGATGTGGT GGGAAAGATT
651 AATGAAAGGT TTATTCGTCT TCTGGCTCTT TGTAAAGATG CTCTCTCAT
701 TGATGACGAG CTCAACATCC TGCCCATCTC CTCCACAGTT GCACCAATGG
751 AGGCTCTGCC TCCCGAGACT CCGGATGAGA GTCTTGGTCC TTCTGATCTG
801 GAGCTGAGGG AGTTGAAGGA GAGCTTGAG GACCCGAGC CTGTGGGTGT
851 GTTGGTGGAC TGGTGAAGA CTCTAGACCA GGCACAAAGT GTCTTGAAT
901 TTATCGAGGG CATCTCTGAA AGACCCCTGA CGATGCTCT TGCATCACA
951 GCTGCTCGAG GACGGGAAAG ATCTCGAGCC CTGGGATTGG CGATTCCCTG
1001 GCGGCTGGCA TTTGGTACT CCAATATCTT TGTACCTCC CCAAGCCCTG
1051 ATAACCTCCA TACTCTGTTT GAATTTGAT TAAAGGATT TGAATGCTCTG
1101 CAATATCAGG AACATCTGCA TTATGAGATT ATCCAGTCT TAATCTCTGA
1151 ATTAAACAAA CGAGTCATCA GACTGAATGT ATTTGAGAAA CACAGGAGAA
1201 CTATTCACTA TATACATCTT GCAGATGCTG TGAAGCTGGG CCAGGCTGAA
1251 CTAGTTGTGA TTGATGAAGC TGCCGCCATC CCGCTCCCTT TGGTGAGAG
1301 CTACTCTTGG CCCTACCTTG TTTTCAAGG ATCCACCATC AATGCTATG
1351 AGGGCACTGG CCGTCACTG TCCCTCAAGC TAATTCAGCA GCTCCCTCAA
1401 CAGAGCGGCC AGAGCCAGGT CAGCAGCACT GCTGGAATA AGACCAGGAC
1451 GACAGCCAGA TTGGCATCAG CGCGGACACT GCATGAGGTT TCCCTCCAGG
1501 AGTCAATCCG ATACGCCCTT GGGGATGAG TGAAGAAGTG GCTGAATGAC
1551 TTGTGTGGCC TGGATGGCT CAGACTCAT CGATAGCTT CAGGCTGCCC
1601 CTTGCCCTGA GCTTGTGAAC TGTACTGTG TAATAGAGAT ACCCTCTTTT
1651 GCTACCACAA GGCCTCTGAA GTTTCTCTCC AACGGCTTAT GGCCTCTAC
1701 GTGGCTTCTG ACTACAAGAA CTCTCCCAAT GATCTCCAGA TGCTCTCCGA
1751 TGCACCTGCT CACAATCTT TCTGCTTCT GCTTCTGTG GCGCCACCC
1801 AGAATGCCCT TCCAGAAGTG CTGCTGTGA TCCAGTGTG CTTTGAAGGG
1851 GAGATTTCTC GCCAGTCCAT CTTGAACAGT CTGTCTGAG GCAAGAAGGC
1901 TTCAGGGGAC CTGATTCAT GGACAGTGTG AGAACACTTC CAAGATCCAG
1951 ACTTGTGGTG TCTGTCTGTG GAAGGGGTG TTGCACTTG TCTTACCCA
2001 GATTATCAAG GGATGGGCTA TGGCAGCGGT GCTCTGAGC GCTGAGAT
2051 GTACTATGAA GGCAGTTTC TTGTCTGGA GGAAGAAGTC CTTGAGACAC
2101 CACAGGAAAT TCACACGTA AGCAGGAGG CTGTGAGCTT GTTGAAGAG
2151 GTCATCACTC CCGGAAAGG CTCTGCTCT TACTGCTCA AATTGAATG
2201 GAGGCTCTGC GAACGCTTG ATTACTGGG TGTTCCTAT GGCTTACCC
2251 CCAGGCTCTC CAAGTCTGG AAACGAGCTG GATTGTTC TGTATTCTG
2301 AGACAGACCC CGAATGACCT GACCGAGAG CACTGTGCA TCAATGCTGA
2351 GACGCTCACT GATAGGATG AGGCTGACA GGAAGGCTG CTTCAGGCT
2401 TCTGGAAGAA TTCCGAGCG CGGTTCTAG CTTGCTCTC CTACCACTTC
2451 AGTACCTCTC CTCTTCTCT GGCTCTGAC ATCATTGCA AGAGGAACAT
2501 GGGGAAGCCA GCCCAGCCTG CCCTGAGCG GAGGAGCTG GAAGCACTCT

```

```
2551 TCTCCCTCTA TGACCTGAAG CGGCTGGAGA TGTATTACCG GAATATGGTG
2601 GACTATCACC TCATCATGGA CATGATCCCG GCCATCTCTC GCATCTATTT
2651 CCTGAACCCG CTGGGGGACC TGGCCCTGTC TGGCCCTCAG TCGGCTCTTC
2701 TCTTGGGGAT TGGCCTGCAG CATAAGTCTG TGGACCACTC GGAAGAGGAG
2751 ATTGAGCTGC CCTGGGGCCA GTTGAATGGA CTTTTCAGCC GGATCATCCG
2801 CAAAGTTGTG AAGCTATTTA ATGAAGTTCA GGAAGAGGCC ATTGAGGAGC
2851 AGATGGTGGC AGCGAAGGAT GTGTCATGG AGCCCAAGAT GAAGACCCCT
2901 AGTGACGACC TAGATGAAGC AGCAAGGAA TTTCAGGAGA AACACAAGAA
2951 GGAAGTGGG AAGCTGAAGA GCATGGACCT CTCTGAATAC ATATCCCTG
3001 GGGACGATGA AGAGTGGAAAT GAAGTTTGA ACAAGCTGG GCCGACGCC
3051 TCGATCATCA CCTGAAAAG TGACAGAAA AGGAAGTTAG AGGCCAACA
3101 AGAACCCAAA CAGAGCAAGA AGTTGAAGA CAGAGAGACA AAGAACAAAA
3151 AAGATATGAA ACTGAAGCCG AAGAAATGT GAGAGAAAC TGGGCACTC
3201 GTGTTGATC ATGGGAAGAT ACTCTACTA ACTGAACCTC CTCTGGCTGG
3251 ACTGTTAAAA GCAACGAGAG GCCCGGCCAC ACTGGAAGC TGGCGCGAA
3301 TTCCGCTCTC GGGCCTGTGT GTCTGTGAG TCAACCTGGC TAAAGCCAGA
3351 GTCACTGCGA AATGGCTCTC TTGAAACTT GATGGCTGG CACTGCCATC
3401 TCTAGAATTG CCACGAGTCT CTCTCTTCTC GCCAGTCCA GGGCCTCTCT
3451 TTCTATAAG TTCATATTTT GCTTTGAGCC AGCTTTTAG TCTATTCCC
3501 ACACATGTGG AAGCCACGTT GCCTCTCGAC CGCTGAGGC CTTAAGTAC
3551 ATCCCTTTCT GTGTGTGCC AGGAGCTGC TCGTGGGCG CTGGGATCT
3601 CTTTGTGGAC TTGTACCTGG AGCAGAGGA ACTCCAGTCC GTCCCGGAT
3651 CCATGGCAGC CCGCGGTTAG GTGCGCCAGG GTTGCTGAT GTTGCTTGT
3701 GCTGTTCCAC TCTTGGCTCC AGCAGACCCA CTGTCCAGA AAGCTGTAT
3751 CCTGTAGTTT ATGTAGAAAT CCACATCTGC GTCTCGAGA CTTGTTTCA
3801 CCATTGGGA AAGATCTTGG GGAAGGCCA CTTTCTCGC AGGGGTGAGG
3851 GGAAGGATAG AGAATCTATT TTTAATAAT AACATTCTAG AATGAAAAAA
3901 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
3951 AAAAAAATA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 102 bp to 3176 bp; peptide length: 1025
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: RGD (966-969)
ATP_GTP_A (284-292)

```
1 MRRKVVWRI RILICNVAE RRSLEFVVV DRKQDVVIL HHMLSKATVK
51 ARPSVLMVKY KELGFSHRK KMRDLQKKI KNGTLMIKQD DPFLFIAAT
101 NIRYCYNET HKILGNTFGM CVLQDFEALT PNLLARTVET VEGGGLVIL
151 LRTNLSLKLQ VVTNDVHSR YRTEAHODVV GRFNERFILS LASCKKCLVI
201 DDOLNILPIS SHVATNEALP PPTPDESIGP SDLELRELKE SLQDTPGVV
251 LVDCCKTLDQ ARAVLRFIDG ISEKTLSTV ALTAARGRCK SAALGLAIAG
301 AVAFGYSNIF VTSPPDNLH TLFETVFKGF DALQYQHLN YEIIQSLNFE
351 FWKAVIRVHV FRRHRTIQY IHPADAVKLG QALVVIDEA AAIPLPLVKS
401 LUGPYLVFNA STINGYEGTG RSLSLKLIQ LROGSAOSQV STTAENKTTT
451 TARLASARLT NDVSLQESIR YAPGDVEMW LNDLCLQCL NITRLVGGCP
501 LPEACELYVV NRDTLFCYHK ASEVFLQRLM ALYVASHYKN SPNDLQMLSD
551 APAHHLFCLL PPVPPTONAL PEVLAVIQVC LEGEISROSI LNSLSRGKKA
601 SGDLIPMTVS EQFQDPDFGG LSGGRYVRIA VHPDYQMGY GSRALOLLQM
651 VYECRFPCLL EKVLTPDPI HTVSSSAVSL LEVITTPRD LPFLLLALNE
701 RPAERLDYLG VSYGLTFRLL KFKRAGEVP VYLRTFPNDL TGEHSCIMLK
751 TLTDEDEADQ GMLAAPMKD FRRFLALLS YQSTFSPSL ALNIIQNRNM
801 GRPAQPALSR EELEALPLPY DLKRLHYSR NHVDVHLND HIPAISRIYF
851 LAHGLGLALS AQSALLGT GLGRKSVQCL EKSELPSGO LAGLPIRIIR
901 KVVKLFNEVO EKAIDQWVA AKDVMKEPM KTLSDOLDEA AKFQEKHKHK
951 EVGKLKSMOL SEYIIRGODE DNVNVLNKG PHASISLSKS DKRRKLEAKQ
1001 EPKQSRKLKN RETNKKDMK LKKKK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZptest3_6c11, frame 3

TREMBL:CEAF3130.4 gene: "F55A12.8"; *Caenorhabditis elegans* cosmid F55A12., N = 1, Score = 2782, P = 1.1e-289

PIR:S55151 probable membrane protein YNL132w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 2549, P = 3.5e-273

SWISSPROT:YXK1_ACHAN HYPOTHETICAL PROTEIN (FRAGMENT)., N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I., N = 1, Score = 2843, P = 3.8e-296

>SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I.
Length = 1,033

HSPs:

Score = 2843 (426.6 bits), Expect = 3.8e-296, P = 3.8e-296
Identities = 576/1033 (55%), Positives = 750/1033 (72%)

Query: 1 MHRKKVDNRIRILIENGVAERQSLFVVVGDGRGQDVVILHMLSKATVKARPSVLNMYK 60
M +K +D+RI LI+NG +E+QRS FVVVGDR +DQVV LH +LS++ V ARP+VLM YK
Sbjct: 1 MPKKALDSRIPTLIRNGCOEQRSFFVVVGDRARDQVNVHLLSQSKVAARPNVLMHYK 60

Query: 61 KEL-GFSSHRKKRRLQKKIKNGTLMIKQDDPFELFIATNIRYCYNETHKILGNTFG 119
K+L GF+SHRKKR +++K+K G + +DPFELF + TNIRYCY E+ KILG T+G
Sbjct: 61 KDLGFTSHRKKRKNKIKKIKNGIRDNSEDPPFELFCSTNIRYCYKESEKILGQTYG 120

Query: 120 MCVLQDFEALTPNLLARTVETVEGGGLVILLATNMSLKQLYTVMDSVRYRTEAHODV 179
M VLQDFEALTPNLLART+ETVEGGG+VV+LL +NSLKQLYT++ND+HSRYRTEAH DV
Sbjct: 121 MVLQDFEALTPNLLARTIETVEGGGIVVLLLNKLSLKQLYTSMOIHRYRTEAHSDV 180

Query: 180 VGRFNERFILSLASCKKCLVIDDQNLIPISSHVATHEALFPQTPDES LGPSDLELRELK 239
RFNERFILSL +C+ CLVIDD+LN+LPIS ++ALPP +++ + ++EL+
Sbjct: 181 TARFNERFILSLGNCENCLVIDELNVLPIGG-KNVKALPPTLEEDN--STQNSIRELQ 237

Query: 240 ESLQOTOPGVVLVDCCKTLQDAKAVLKFIEGISEKTLRSTVALTAARGRGSAAALGLAIA 299
ESL + P G LV XTLQDA+AVL F+E I EK+L+ TV+LTA RGRGSAALGLAIA
Sbjct: 238 ESLGEDHPAGALGVYKTLQDAKAVLTFVESIVKSLKGTVSLTAGRGRGSAAALGLAIA 297

Query: 300 GAVAFGYSNIFVTSPPDNHLTLFETFKGFDALQYOENLDYELIQSLNPEFNKAVIRVN 359
A+A GYSNIF+TSPSP+NL TLFEF+FKGFDAL Y+EH+DY+IIOS NP ++ A++RVN
Sbjct: 298 AAIANGYSNIFITSPSPENKTLTFETFKGFDALNVEEHVDYDIQSTNPATNMAIVRVN 357

Query: 360 VFREHRTIQIYHPADAVKLGQALVVIDEAAAIPLVKSLLGPYLVFNASTINGEYGT 419
+FR+HRTIQIYI P D+ LGQALVVIDEAAAIPLPLV+ L+GPYLVFNASTINGEYGT
Sbjct: 358 IFRDHRTIQIYISPEDSNVQALVVIDEAAAIPLPLVRLIGPYLVFNASTINGEYGT 417

Query: 420 GRSLSLKLQQLRQOSQSGVSTTAENKTTTARLASARTLHEVSLQESIRYAPGDAVEK 479
GRSLSLKL+QQLR+OS S + NK+ + + + S RTL E+SL E IRYA GD +E
Sbjct: 418 GRSLSLKLQQLREQSRI--YSGSGNKSQDSQSHI-SGRTLKEISLDEPIRYANGDRIEL 474

Query: 480 WLNDLCLDCIN-ITRIVS-GCLPEACELYVNRDTLFCYHKASEVFLQRLMALYVASH 537
WLN LCLCD + ++R+ + G P P C LY V+RDTLF YH SE FLQR+M+LYVASH
Sbjct: 475 WLNKLLCLDAASVSRNATOGFPHPSECSLYRVRDTLFSYHPISEAFLQRMMSLYVASH 534

Query: 538 YKNSPNDLQMSDAPAKHLFCLLPVPPTQNALPEVLAVIQVCEGEISRQSIINSLSRG 597
YKNSPNDLQ++SDAPAH LF LLPPV LP+ + VIO+ LEG ISR+SI+INSLSRG
Sbjct: 535 YKNSPNDLQMSDAPAKHLFVLLPPVDLKNPLPDPICVQLALEGISRESINSLSRG 594

Query: 598 KASGDLPWTVSEOFQDPDGGLSGGRVRIAVHPDYQMGYGSRALQLQNYEGRFP 657
++A GDLPW +S+QFQD +F L G R+VRIAV P++ HGYG+RA+QLL Y+EG+P
Sbjct: 595 QRAGGDLPWLSQOQFQDFAALGGARIVRIAVSEPHVMGYGTRAMQLLHEYFEGKFI 654

Query: 658 CLKKEVLETPOEINTVSSEAV---SLLEEVIIPR--KOLPPLLLKNERPAERLQYGV 712
E+ + + E + +L E I R K +PPLLLKLE E L Y+GVS
Sbjct: 655 SASSEFKAVKHSKRIQDEIENALQTEKIHVRDARTMPPLLLKLELOPEPLHYGV 714

Query: 713 YGLTPRLLPKPKRAGFVPPVLRQTPMOLTGENSCIMLKTITDEEADQCGWLAAPKDFR 772
YGLTP L KPKR G+ P+YLRQTPMDLTGEN+CM+L L D ML AF ++F
Sbjct: 715 YGLTPSLQPKKREGYCPYLROTANDLTGENTCVNLRVLEGRDSE---VLGAFAQNFY 770

Query: 773 RRFLLALLSYOFSTSPSLALNLIQNRNNGKP---AQPALSREELEALFLPYDLKRLNY 828
RRFL+LL YOF F+ AL+++ NG + L+ EE+ +F YDLKRL Y
Sbjct: 771 RRFLLSLGYQREFAAITALSVLDACNNGTEYVNVNSTSKLTNEIINNPFESYDLKRLSY 830

[illegible]

Pedant information for DKFZphtes3 6c11, frame 3

Report for DKFZphtes3_6c11.3

```

[LENGTH]      1025
[MW]           115704.57
[pI]           8.50
[UNIMOL]       P1355151 probable membrane protein YNL132w - yeast (Saccharomyces cerevisiae)
[0]
[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YNL132w] 0.0
[C]           C: general function prediction [H. influenzae, H1254] 2e-05
[PROSITE]
[PROSITE]      ATP_GTP_A 1
[PROSITE]      RGD 1
[EW]           Alpha Beta
[EW]           LOW COMPLEXITY 11.80 %

```

```

SEQ      MHKKKKVQNRIRITLITENQVAERRSFLPVVQDRGQDVVLIHMLLSKATVKARPSVLVMCKY
SEG
PRD      ccccccehhhhhhhhccccccccccccccccccccceehhhhhhhhhccccceehhh

SEQ      KELGFSHRRKMMRGLQKKIKMGLTIRKODOPFELFIATNIRYCYVTHCKIKLQTFGN
SEG
PRD      hhcccccehhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ      CVLQDFEALPNILARTVETVGGGLVILRLTHNSLKQLYITVMDVHSRYTRTEAHQDVV
SEG
PRD      eehhhhhcccccehhhhhhhhhhcccccccccccccehhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      GRNEFRITLSLASCCKVLVIDQMLPLPISSHVATNEALPOTPOESLQPSDLEALKEK
SEG
PRD      hhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccceehhhhhhhhh

SEQ      SLQOTPGVGVLDCKRTLDAQAVLKFIQSEKTLRSTVATAARGKSAAGLALGA
SEG
PRD      hhccccccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      AVAGYSNIPVTPSPDNLTFFEFKFGDAPQYQHLOYEIIQSLNPFNKAVIRVNV
SEG
PRD      hhhhhccccccccccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      FREDHQTIQITPADAVYQLQAEVLVIDEAAVPLPLVKSGLPVVFMASTINGEVGT
SEG
PRD      hhhhhhhhhccccccccccccccccccccceehhhhhccccchhhhhhhcccccccccccccc

SEQ      RSLSLIKQLQROGSAOSVSTAEHTTARLASARVSLVQLQESIYRAPGDAVEKW
SEG
PRD      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
ccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      LNDLCLDCLNITRVSQGLPEACELEYVYWRDTLCFYHAEVFLQRLMALVYASHYKW
SEG
PRD      hhhhhhhccccccccccccccccccccccccccccccccccccceehhhhhhhhhhhhhhhhhcc

SEQ      SPNQLQMLSDAPAHLLPCLPLVPVPTQNALPEVLAVIQVCELETSRQSLNLSRGGKA
SEG
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhccccchhhhhhhhhcccc

SEQ      SGBLIPMTVSFEQPOFGGLGGVRVIRIVHFDYQMGYGRALQLQMQYEGFPCLE
SEG
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      EKVLEPTQPIHTVSVSEAVSLSEVITPRKDLPLLLKNERPAERDLYVGSYGLCPH

```

WO 01/12659

PCT/IB90/01496

```
SEG .....XXXXXXXXX.....
PRD hhhhhccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccchhh
SEQ KFWKRGFVPPVYLAQTWDLTGKSCIMLKTITDEDEADQGGWLAAPKDFRRFLALLS
SEG .....
PRD hhhhhccccccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ YQFSTFSPLALWIIQNRHMKPAQPALSRELEALFLPYDLKRLEMYSRNMVDVHLINO
SEG .....
PRD hhhhhccccccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ MIPATSRITYFLNQLGDLALSAQSAALLGIGLQHKSVQLEKEIETPSQQLMGLFWRIIR
SEG .....XXXXXXXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ KVVKLFNEVQEKAIIEQWVAADVMEPTMTLSDDLDEAAKEFOEKHKKEVGKLSMDL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ SEYITRGDDEEMNEVLNKGPHASIIISLKSOKRKLKAEQEPKQSKLKKHRETNKKDKK
SEG .....XXXXXXXXXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ LKRKK
SEG KKKKK
PRD hhccc
```

Prosite for DKFZphtes3_6c11.3

```
P500016 966->969 RGD PDOC00016
P500017 284->292 ATP_GTP_A PDOC00017
```

(No Pfam data available for DKFZphtes3_6c11.3)

DKFZphtes3_6d16

group: testes derived

DKFZphtes3_6d16 encodes a novel 695 amino acid protein nearly identical to a sequence from human PAC Clone WUGSC:H_QJ1185107.2.

The cDNA is different to the proposed gene model: it contains additional exons.
No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_QJ1185107.2, differences to genmodel

differences to genmodel of WUGSC:H_QJ1185107.2 two exons skippt.

Sequenced by BMFZ

Locus: /map="7q11.23-q21"

Insert length: 4572 bp

Poly A stretch at pos. 4540, polyadenylation signal at pos. 4520

```
1 GGGGGCGCTA GCTTCGAGT C'CCCGCGG CACCTAGCC GCTCCTAGC
51 GGGGGCGCGC TCGCTCCTAC GCTTAAATG ACCAATGTG GTTTCAGTG
101 GAATAAATGG CGTCCAAAGT CACAGATGCT ATAGTCTGGT ATCAAAAGAA
151 GATTGGAGCA TATGATCAAG AAATATGGGA AAATCTGTT GAACAGAGAG
201 AAATCARGGG CTAGAGATG AATCAAGA AATCAGACA TGTGAAGCA
251 GACCTCATAG ATGTTGATCT TGTGAAGGG TCTGCATTG CAAGGCCAAA
301 GCTGAAAGT CTTTGGACTT CTCTGACCAG AAAGGAANT GTTCGAGTTG
351 TATTTTTCCT CTTTTCCTTC CGGTGGTGGT TACAAGTAC ATCAAAAGTC
401 ATCTTTTCTT GCGTCTCTGT CTTTATCTT CTTCAGTTG CTGAATAGT
451 ATTTATCTGC TCCACTTCTA GCGCCACAG CATACTCTG ACAGAGGTGA
501 TTGGGCGCAT ATGGCTGATG CTGCTCCTGG GAACGTGGA TTGCCAGATT
551 GTTTCACACA GAACACCCAA ACCTCTCTTA AGTACAGGG GTAAAGAGAG
601 AAGCAAAATTA ACAAAGGAG CCCATTTGGA AGTACATGG GAGGAGATG
651 GTTCTAGTAC CACAGATAAC ACACAAGAG GACCACTTCA GAACCCGGT
701 ACAAGCACCT CTCACAGCGT TGGCACTGTC TTCAGAGATC TCTGGCATGC
751 TGCCTTCTTT TTATCAGGAT CAAAGAAAGC AAGAATTCA ATTGATAAAT
801 CACTGAAGC TGCAATGGC TATGTATGCC TTGATGGGA CAAGACTGTT
851 AAAGCGGGT AGATGGAAT ACAAAACCAT GAACCTCAGT GTGAAACTAT
901 TCGACCGAGG GAGACAGCCT GGAACACAGG AACACTGAGG AATGGTCTTA
951 GCAAGATATC CCAAGGACCA ATAAACAATG TCTCTGATGA AGTCTCCAGT
1001 GAGCAAGGCT CTGAACAGG ATACTATTA COTGCTCATG TGGCAGGAC
1051 TTCTGAAGGT GTTCTTGGGA ATGAAAGTC ACACCATTA AGAAACATT
1101 ACCCTAATGA GGACGCCCTT AAATCGGGTA CTAGTTGAG CTCTCGCTGT
1151 TCAAGTTCCA GACAGGATTC TGAGAGTGCA AGGCCAGAA CTGAACACGA
1201 AGATGTGTGA TGGGAAGACT TTTTCAATG TGCAGATGC CATTCACTTT
1251 GTACCAGTGA CACAGATGTG GAAATCATC AGATTATCC ATGTGTGAAA
1301 AAAGAATATA GAGATGACCC TTTTCATCAG AGTCATTGC CTTGGCTCCA
1351 TATGTCCCA CAGAGATTAG AAAAATAAG TGCTATAGTA TGGGAAGTGA
1401 ATGATGTGTA GAAGCAGAC ATGCTCTGAC TTGAATCAG TGTATGATA
1451 ATGAACAGAG TGAACAGCCA TATACAGGA ATAGGATACC AGATTTTGG
1501 AATGCAAGTC TCTCTATAC TGGGTTTAA TCCATTGTT TTCCGACTTT
1551 CTGAAGCTAC AGACTTGGAA CAATCTCAG CACATTCTGC TTGAGACTTT
1601 TATGTGATG CATTGGTCT TAATGAAGAT GTCATAGTTC TTCTATAGT
1651 TATAATAAG TTTGTGGTTC GCGTCTCTC TGTGTGATT TTCTTTTTTT
1701 TGCTCTGTGT AGCAGAAAGA ACTTATAAAC AGCGATTACT TTTTGCAAAA
1751 CTCTTTGGAC ATTTAAGATC TGCAAGGAGG GCTCGAAAT CTGAGGTTC
1801 TCACTTCCGG TGAAGAAAG TACAGATAT AAAAATGCG CTATCTCTCC
1851 GTTCTATCT TAAGCGTCTG GGTCTCAGC GATCAGTTGA TGTAAAGTCT
1901 TCATCTGCTT TCTTATTGAC TATCTCAGT GTATTATCT GTTGTGCCCA
1951 GATAAACCTC TACTTGAATA TGGAGAAAA ACCTACAAA AAGGAGGAGC
2001 TACACTACTG GAATAATGTT TTAATACTGG CTACTAACT GCTAAGGAG
2051 TTGACAGCTC CTTTATGATT ATATGGGCTT ACAATGAAT CGCTGCTTTA
2101 TAACATCAAC CAGGTGTGTA TCTGTGAGC TGTTTCTGTT GTTATCAGTG
2151 ACTTGCTTGG ATTTAATTTA AAGCTATGGA AGATTATGTC ATGACAATTC
2201 AAAGAAAGCA AGATCTAGCC TCTTTTCCAG AATAAGATA CTGACTAAGC
2251 TGGCTGAAG CTTCTCAGCT ATCTTTGCT TCAAGAGTCT CAGCTAGGGA
2301 GTTGAAGTGT TTACATCAGA CTGCTTGTG CAATTCTTAT ATTTATTTTA
2351 CTGGTCACT TTTTITTTACA TTTATTTTAT TCTTATATT TTTATTTTTA
2401 AGCATTTGAT TACTTATGTT TGAAGAGGT GATGAAGTCT ATATCCAGAT
2451 ACTTGAGATC CTGGTAAATG GTCATAATA ATTTGCCAAA TAACAATTTG
2501 TGAATAATGA AGCCATTGCT CAGCACCGTT TCTCCATCAA TGCCGTGAAC
2551 TTGCTTACT TGAGAAAAA TTCTTTAACT TTGCAATATT GCATTGAAC
2601 CAGCTATACA CATAAAAAAT TTCTTTTGGT AATCAAGAT CCACTCAGGG
```



```
2651 TTTCTCTTGA ATTATTTGGG AACAATGCCA GGATCCAAC TGATTAAGTT
2701 ACAGTTTAAAG CACCCCTCAG TATTAATATA TACGGTATTA TATACAGGT
2751 CACAAGTGGC TCTTTGATCA TAAAACTTGT AATGAGGCAA TAATTCTAAA
2801 TGGTTACCAT ACTGTAAGAT ATTTTGATAA AAATTAACCTA GTAATCTTG
2851 TATTTATTGG AAACACTGGG CTGTTTGAC AGCTCCAAC GTGCATGCTC
2901 AAAATGTGCA CTTTITAAAA TTGTTACTTT TAATGGGTAT CTTTATATGG
2951 GATCTGTTAT AGTATACTAG GGCATGATAT GGTATCCTTT TGAGTGAGGT
3001 ATATACCTAT CTCACAAGTG AAGTGCTTAC TGATATTACT AAAGTACATT
3051 ATGTTTACTC AAGTAAATAA TTTTCTCCCC ATGGTACACT CTAGTGTAGG
3101 CTATTCATAC CACACTGAAA TGAAACACTG AAGATTAAGG CTAGCAACCA
3151 ATAAAAATTT TCTCTAATTG CTAGTTGTAA AACTGTATCC AAATTTTCAG
3201 AAAAGACAGC TTCAGCTTGC AAATTCATAT CTCTAAACCT ATCTGGTGCA
3251 TTCTCCCCAC CCCACCCCCA TTATATAAGG GCTATTTTAG ATGCTTTTAA
3301 CTTCCCAAC AAATTAATTTG CCAAGTGTCC AATGAGAACT TATCATGTG
3351 CTGTGTTAGG TAAATCGGCC AAATATGATA GTGCTTTACA TTGGGCTTG
3401 ATTTTAAGTT GTTATATTG TACAATCGAG TATTTAGAA ATTACATGAA
3451 ACATGAACA GTTTTTGCAA TTTTITTTAA ACTGGGCATC TGGTTCTTAA
3501 AAATTTATTT AAAACAATCT AGAATTTTCT TGGTGCAAG TGTATCATGT
3551 GGAATATCTT CATATTTTAA CCAATTTTAA AGAATCTTAA GAGGATTAAT
3601 TGTAAATAAT TTATTTGATT GTGCGAGTTC TAATCCCTAA ATCATAATCT
3651 TAAAAATCAG AATGTGTGGA GAACAGAGCC ATGTGATATC ACTTTGCTCT
3701 TACCATTCTT TTGATCAGC CTCAAATTCG CTTCAATGTG TAGTATGTTT
3751 TTTCTTTCTA TGAANAACAA CAGAAAGCAT TTCATTTTAT TTGCTATGT
3801 TCAAAATATG TTAATAATGA CCAAAAGTGA TTCTGAGTTT TTCAAGGAA
3851 TGTAAATCTG GAGCTTTAAG AACATACCTA GTTTCTCATG TGAACACTTA
3901 GGGCTTGTCT GATGTTTTC CTCTCATATG TGTCTAATGT TGAGGTGTTT
3951 TTGAGAAATT ATGTTTATA AACTTTTICA ATATAAGGTA CATGCCATATA
4001 CAGAACTTAA CATTITGCAC AGAATATATC AAATATATTT TGAGAAAAAA
4051 AGTACGGCAT GAGTTCTCTT AGGAATAAAA GATGAACCTA TTGATCTCA
4101 CAAAAAATCT TATTTCAGAA TGAATAATTT TTGAGAAATA GTAGCTGAGT
4151 ATACTGGTTT AAGAAATGC TTGTTTITAA TTGAGGTTAA CTTAGAGTTG
4201 GGAGTGATTT TATTAAGTAC AGTATACCTC TCAACAGTTT ATAAATAATA
4251 TGTGAAATTA TGTCAGTGTG GGCAGCAGTA GAATCTAATA AGGAAATGT
4301 CATGTTAAGC AATTTCAGAA CATTACTGTA ACTATTTTCA AAGCAGAAAA
4351 ATTGACATTT CTGCTTTTAA GAATACCATG AATGTAGAAA ATTGAAGAAA
4401 ATTGTAAAA ATCACATAAT ATAGAAATGG CAGTTCAAAG AGAATTTGG
4451 CAGATGTTGT GTGTGAACGT TGTTTCTTCT GCCACATGTG TTGATTTTGA
4501 AAGTTTACA GTAAAGTTAA AATAAAACAT TCTGTGACTG AAAAAAATA
4551 AAAAAAATA AAAAAAATA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 2191 bp: peptide length: 695
Category: known protein
Classification: unclassified
Prosite motifs: CYTOCHROME_C (375-381)

```
1 MASXVTDIV WYQKIGAYD QQWEKSVEQ REIKGLRNP KTAHVKPD
51 IDVDLVGSA FAKAKPESFW TSLTRKGIV VVFPFFFRN WLQVTSKVIF
101 FWLLVLLQ VAAIVLFCST SSPHSIPLTE VIGPIWMLL LQTVHCQIVS
151 TRTPPLST QERRRRLKAR AALLEVHREO DGSSTTMDT EZAVQHQDTS
201 TSHSVCTVF DIRMHAATFLS GSRAKNSID KSTETDQIVY SLDGKTVFS
251 GEDGQNHPE QCTIRPEET AMNTGLENG PSKOTQRTI NVSDEVSEEE
301 GPETGYSLRR HVDTSEGV L RNRKSHHYK HYPNEDAPKS QTSKSSRCS
351 SRQSEARP ESDTEVME DLLHCAECHS SCTSETDVEN HOINPCVKEE
401 YRDPFHQSH LPMHSHHPG LKLSIAIWE GNOCKKADS VLEISQINW
451 RVNSHIPGIG YQIFGNVSL ILGLTPFVFR LSQATOLEQL TAHSASELYV
501 IAFGSNEDVI VLSMVIIIFV VRVSLVWIFF FLICVAERTY KORLLFARLF
551 GHLTSARRAR KSEVPHRLR KVNQIVKMLS LSTLARRGP QRSVDIVS
601 AFLITISVF ICACIHLVL RDKKPNKKE ELTLVNVKL LATKLKELD
651 SPFRLVGLTM NPLLYNITQV VILSAVSGVI SLLGFNLKL WKIKS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6d16, frame 2

PIR:S38170 SRP40 protein - yeast (Saccharomyces cerevisiae), M = 1,
Score = 100, P = 0.00

TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185107.2": Homo sapiens PAC clone
DJ1185107 from 7q11.23-q21, complete sequence., N = 2, Score = 2693, P
= 0

>TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185107.2": Homo sapiens PAC clone
DJ1185107 from 7q11.23-q21, Complete sequence.
Length = 588

HSPs:

Score = 2693 (404.1 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 510/515 (99%), Positives = 512/515 (99%)

Query: 35 GLRNPKPKTAHVRLDIDVLRGSAFAKAPESPMTSLTRKGIVRVVFVFFFRWMLQV 94
Sbjct: 1 GLRNPKPKTAHVRLDIDVLRGSAFAKAPESPMTSLTRKGIVRVVFVFFFRWMLQV 60
Query: 95 TSKVIFFWLLVLLQVAAIVLFCSTSSPHSIPLTVIGPIWMLLLGTVMCQIVSTRTP 154
Sbjct: 61 TSKVIFFWLLVLLQVAAIVLFCSTSSPHSIPLTVIGPIWMLLLGTVMCQIVSTRTP 120
Query: 155 KPPLSTGGKRRRLRKAHLEVHREGDGSSTTONTQDGAQVQNGTSTSHSVGTVFRDLWH 214
Sbjct: 121 KPPLSTGGKRRRLRKAHLEVHREGDGSSTTONTQDGAQVQNGTSTSHSVGTVFRDLWH 180
Query: 215 AAFPLSGSKKAKNSIDKSTETONGVVSLOGKRTVKSDEGQIQNHPOCETIRPEETAMNT 274
Sbjct: 181 AAFPLSGSKKAKNSIDKSTETONGVVSLOGKRTVKSDEGQIQNHPOCETIRPEETAMNT 240
Query: 275 GTLRNGPSKDTORTITNVSEVSSEEGPETGYSLRHHVORTSEGLVRNRKSHHYKHYPN 334
Sbjct: 241 GTLRNGPSKDTORTITNVSEVSSEEGPETGYSLRHHVORTSEGLVRNRKSHHYKHYPN 300
Query: 335 EDAPKSGTSCSSRCSSRRQDSARPESETEDVWEDLLHCAECHSSCTSETDVENHQIN 394
Sbjct: 301 EDAPKSGTSCSSRCSSRRQDSARPESETEDVWEDLLHCAECHSSCTSETDVENHQIN 360
Query: 395 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADNSVLEISGHIMNRVNS 454
Sbjct: 361 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADNSVLEISGHIMNRVNS 420
Query: 455 HIPGIGYQIFGNAVSLILGLTPFVRLSQATDLEOLTANSASELYVIAFGSHEDVIVLSM 514
Sbjct: 421 HIPGIGYQIFGNAVSLILGLTPFVRLSQATDLEOLTANSASELYVIAFGSHEDVIVLSM 480
Query: 515 VIISFVVRVSLWIFFFLLCAERTYKORLLFAKL 549
Sbjct: 481 VIISFVVRVSLWIFFFLLCAERTYKQINLYLKM 515
Score = 409 (61.4 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 92/115 (80%), Positives = 98/115 (85%)
Query: 595 DVIVS---AFLLTISVVF-----CCA-----QINLYLKMKEKPNKKEELTLVNNVLK 640
Sbjct: 474 DVIVLSHVIISFVVRVSLWIFFFLLCAERTYKQINLYLKMKEKPNKKEELTLVNNVLK 533
Query: 641 LATKLLKELDSPFRRLYGLTNPLLYNITQVILSAVSGVISDLGLFNKLKMKIKS 695
Sbjct: 534 LATKLLKELDSPFRRLYGLTNPLLYNITQVILSAVSGVISDLGLFNKLKMKIKS 588

Pedant information for DKFZphtes3_6d16, frame 2

Report for DKFZphtes3_6d16.2

(LENGTH) 695
(MW) 78466.68
(pI) 9.30
(HOMOL) TREMBL:AC004990.1 gene: "WUGSC:H_DJ1185107.2": Homo sapiens PAC clone DJ1185107
from 7q11.23-q21, complete sequence. 0.0

```
(PROSITE)      CYTOCHROME_C      1
(KW)            TRANSMEMBRANE      6
(KW)            LOW_COMPLEXITY      5.32 1

SEQ      MASKVTDAIVVYQKRIAYDQDIWEKSVEQREIKGLRNKPKKTAHVKFDLIDVLRGSA
SEG      .....
PRD      cccceeehhhhhhcccccchhhhhhhhhhhhhhhcccccceccccceccccccch
MEM      .....

SEQ      FAKAKPSPMTSLTRKGIVRVVFVFFFRMWLQVTSKVIFFMLLVLYLQVAAIILFCST
SEG      .....
PRD      hhhccccccccccccceccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM      .....

SEQ      SSPHSITPLEVIGPIWMLLLGTVHCQIVSTRTPPLSTGGKRRRLRKAHLEVRREG
SEG      .....
PRD      cccccceeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM      .....

SEQ      DGSSTDTMTQEGAVQNGTSTSHSVGTFRDLWAAFFLSSGSKAKNSIDKSTETDNGTV
SEG      .....
PRD      cccccceccccccccccccchhhhhhhhhhhhhhhcccccchhhhhcccccceccc
MEM      .....

SEQ      SLDGKKTVSGEDGIQWHPQETIRPEETAHNTGTLRNGPSKDTQRTITNVSDVSSEE
SEG      .....
PRD      ccccccecccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      GPETGYSLARHVORTSEGVLNRKSHHYKHYFWEDAPKSGTSCSSRCSSRQDSASAP
SEG      .....
PRD      cccccceccccccccchhhhhccccccecccccccccccccccccccccccccc
MEM      .....

SEQ      ESETEDVLWEDLLHCAECHSSCTSETDVENHQINPCVKEVRDDPFHQSHLFWLHSSHPG
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      LEKISAIWEGNCKKADMSVLEISGHNHVRVNSHIGICVQIFGNVSLILGLTPFVFR
SEG      .....
PRD      cccccceccccccccceeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      LSQATDLEQLTAHSASELYIAFGSNEQIVLSHVIISFVVRVSLVWIFFLLCVAERTY
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      KQRLFLAKLFGHLTSARRARKSEVPHFLKVKQNIKMWLSRLSYLKRGPQRSVDIVSS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      AFLLTISVVFICCAQIMVLYLNEKKPKWKEELTVNNVLRATKLKELDSPFLYGLTM
SEG      .....
PRD      eeeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      NPLLYNITQVILSAVSGVISDLGFWLWKIKS
SEG      .....
PRD      cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....


```

Prosite for DKFzphes3_6d16.2

PS00190 375->381 CYTOCHROME_C P00C00169

(No Pfam data available for DKFzphes3_6d16.2)

WO 01/12659

PCT/IB00/01496

DKFZphtes3_72k11

group: testes derived
DKFZphtes3_72k11 encodes a novel 233 amino acid protein with similarity to S.pombe
hypothetical repeat-containing protein.

The novel protein contains 5 leucine zippers and a microbodies C-terminal targeting signal (S-K-L) signature. This sequence is responsible for transport of proteins from free polysomes into the microbodies.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.pombe hypothetical repeat-containing protein

complete cDNA, complete cds, 6 EST hits (3 from testis derived libraries)

Sequenced by DKFZ

Locus: unknown

Insert length: 1134 bp

Poly A stretch at pos. 1124, polyadenylation signal at pos. 1088

```
1 AACCTTTCAA GTGCCCCCTC CTTTCCTTAA AGTCTTTTAT AGGGGTCCCC
51 TTCTTGCCCA TCTCCATCCT GTGAGTCAGG ACTGAAAGGG CACAGACAGG
101 TCACTGCCAG CATTTGTTGG GCAAGCCTGC AAGCAGCAT CACTGGGAT
151 CTGACATGAC AATGCGGCC TGCCTCCTCT CAGGGCTACA GGACTTACCC
201 CAGTGGGAAG CAGCTAAGCA GGTCTGACCA GCCGACCTGG ACCTGGCCAA
251 GGTCCCTGTC ATCCCTCATG CCCACCCCGC CATTCCGGCT GATAAGGAAG
301 ATGTTTCTCT TCAAGGTGAG CAGATGGATG GGGCTTCTCT GCTTCCGTC
351 CTGCGGCCCA TCCTCTCCCA GTATTGCGCA GAAGAACTA ATGCACAGC
401 TGCAGGAGGA AAAGGCTTTT CGCGAAGAGA TGAATAATTT TCGTGAAGAA
451 ATAGAGGACT TCAGGGAAGA GATGTGGACT TTCCGAGCCA AGATCCATGC
501 TTCCCGGGGC CAGATCCTGG GTTTTGGGA AGAGGAGAGA CTTTCTGGG
551 AAGAGGAGAA AACCTCTGG AAAGAGGAAA AATCCTTCTG GGAATGGAA
601 AAGTCTTTCA GGAAGGAAGA GAAACTTTC TGGAAAAAGT ACCGACCTTT
651 CTGGAAGGAG GATAAGGCTT TGTGGAAGA GGACATGCC TTATGGGAAA
701 GAGACCGGAA CTTCTTTCAG GAGGCAAGG CCTGTGGGA GGAAGAAAG
751 GCCCTGTGGG TAGAGGAAG AGCCCTCCTT GAGGGGGA GAAGCCTGTG
801 GGAAGATAAA ACCTCCCTCT GGGGGAAGA GAATGCCCTC TGGGAGGAAG
851 AGAGGGCCTT CTGATGGAG AACATGGCC ACCTTCCGG AGAGCAGATG
901 CTCGAAGATG GCGCCCAAG GGCACAGAG GGGCAGCCT TCTTGGCCTT
951 CTCGGAGGCG AGGGCGTAGC CAGCATGAG GTGCAGGCC CTGTGGTCCA
1001 GACTCCCTGT GGTGGGATT CAAGTCAGG GTGAGCCAT GTCTGGAGA
1051 AAATACAGAC TCATTGGTCT CTTGCTTTG AAAGATCAA TAAAGTCTG
1101 AGGCAAGGTT TGGAAACCA ACTTAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 268 bp to 966 bp; peptide length: 233
Category: similarity to known protein
Prosite motifs: MICROBODIES_CTER (231-234)
LEUCINE_ZIPPER (142-164)
LEUCINE_ZIPPER (149-171)
LEUCINE_ZIPPER (156-178)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (170-192)
LEUCINE_ZIPPER (170-192)

DKFZp372k15

group: cell structure and motility

DKFZp372k15 encodes a novel 188 amino acid protein with strong similarity to Rattus norvegicus actin-filament binding protein Frabin.

FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for facio-genital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

strong similarity to actin-filament binding protein Frabin

2 EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1845 bp

Poly A stretch at pos. 1835, polyadenylation signal at pos. 1816

```
1 GTGATGGAGA GTGCTGTTAT GATAGATGAA TTAGGAAAG CCTCTTTGGA
51 GATGTGATAC CTGAACAGAA CCCCGAATGA TAAGAAGAAA TACCAGTGT
101 TTAGGAGAGA TTGTCTTAAG CAGAGAAGAC CAGCTGCAAA GACCCGAGA
151 CACATACACT TCGTTATTAA GAATGGCAGC AGCAGGAGCT ATGGCAAGAA
201 CACAGTGAAT TTTCCCTTGA GTGTGTGAGG AAGCCCTCAG AGTTTGTGAC
251 TGACTTTGAG AGGTTCTAGT GGAGGGGATC AGAGTGGAAA CAAAGAGACC
301 AGTTAAAGAG GTATGGCAGC ATGAATAAAA AGCTTTTGAG AGTATTCTAT
351 ATGCTTTCCA AATAAAAAAC TCTTTGGCTC ATAATTGCTT CATATAATTA
401 GGACTGGCTA CACTGTACTA TTTAAAAATG TTAAGAAACA TCAATAAGTA
451 AAAATGTTAG GAAGAGATGA TAAATACGTA AGTATTATAT CTAACTAAGT
501 CTCTACTACG TAGTCACATT ATTAACAGCT CCAAGGATCA AGAANAAGTA
551 AGCGTTGAAA AATAAATAAA TAAGTTATAA ATAAATAAAA CAGCCCAAGG
601 AAATGTTCCA GTCCCATAG GTAGACTCGG GGTCTCTTTC TTTATTTAAA
651 TCTTTATTTA AATGTGGATA GCATCCCAAG AGACTTGGGT CTACACTAAG
701 AATATTCAAA TCCATGTTTC TGAAGCAGCT AGAGATAGAA AAAAAGTA
751 GCGAATATCC CTTTTCAACT GGAATAAAGT TGCTTTAATT CTAGACTTTT
801 TCCATACCAA TGTTTTCATG CTTCCTTTGT ATTTTATCTT TTAGCTCAAT
851 ATCAAAATTAT AGTGATTGGA AGAAAGAGTC TGCTGTGAAC CTAAATGCTC
901 CTGAACCCCC AGCAGGCGCT GCAATTCAGC CCACACCTCA ACAAAAATCT
951 CTCTCCAGCC ACTTGCCACA GAGGCGAGGA AATCAGACAG ATAAGACTCA
1001 GGGTGCCAGC ACTTGTGTGG CCAACGGTGT AATGGCAGCA CAAAACGAGA
1051 TGGAAATGGA GGAGGAGAAA GCTGCGACTC TTAGCTCAGA TACTTGTATT
1101 CAGCTTTGTC AACCTTTGCT GATACCGCAC ATAGTGAATG GAGAAGAGA
1151 TGAAACTGCC ACAGCTCCTG CATCACCCAC AACAGATAGC TTGATGGAAA
1201 ATGCTTTCTGA CAGTAGCTAC AGGACTCCAG GCATAGGCCC AGTGCTGCCC
1251 CTAGAGAGAA GAGGGGAGCA AACAGAAACC AAGGTCAAG AGAGGAGAAA
1301 TGGGAAAGCC CCTCTGGAAC TGGAGCAGCT GGACAGCAGC CATGAGATGA
1351 AGGTAGAGCA TGAGACTAGC TCATGAGCAG GGAAGACCTT GCCTATTGCA
1401 TTGTTGCTCT AAAACTCTTT ATTTATTGCA CCGCTGAATT GTATGAATCA
1451 GATCACCCAC ACTGGCAGTT AAGCATTTT CAGCTCTGCG CTCTGATTA
1501 GCATTTCCCC TATGCTCTAA GCAGATATTT CACTTTTCTT TTTCATGTAG
1551 TTCTGTGTTA TATCTCTGTT GTAATTTTCA GAGTCAGAAC AGTGTGGAAA
1601 CTTTAATATA GGAATCCAC AAGTGATATG TTTTACATA GAAGAAAAAT
1651 GTTCTGTTGT GCTTAGATG TTGGTGTGCT ATCCCTATA CTAGGGGCC
1701 AAGCAAGAGC AAATGTGATA ATCTTTGTTG TTAGAGGTTT TCTAATAGAA
1751 TAAATAGGCC TATAAGATGA ACTTGCACCT AGTAATGTTT ACTTTTAGAG
1801 ACATGAATAT GGAAGTATTA AATTATTCAA CAGATAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

98334590:
Frabin, a novel FGD1-related actin filament-binding protein capable of
changing cell shape
and activating c-Jun N-terminal kinase.

Peptide information for frame 3

ORF from 810 bp to 1373 bp; peptide length: 188
Category: similarity to known protein
Classification: Cell structure/motility
1 MFSCFLCLIS FSSLSNYSYL KESAVNLMA PRTPGRHGLT TTPQKLLSQ
51 HLPQHQGNDT DKTGAGTCV AHGVNAAQHQ MECIEEKAAT LSSDTSIQAS
101 EPLLDTHIVN GERDETATAP ASPTDSCDG NASDSSYRTF GIGVPLPLEE
151 RGAETETKVO ERENGESPLE LQLOQHHEM KVENETSS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72k15, frame 3

TREMBL:AF038388.1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds..
N = 1, Score = 428, P = 1.8e-39

>TREMBL:AF038388.1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds.
Length = 766

HSPs:

Score = 428 (64.2 bits), Expect = 1.8e-39, P = 1.8e-39
Identities = 90/174 (51%), Positives = 115/174 (66%)

Query: 12 SLSLNYSDLKESAVNLMA PRTPGRHGLT TTPQKLLSQ HLPQHQGNDT DKTGAGTCVA 71
S LQ+Y+D+R+S +KLM P+TP +HGLT+T QKL S PQ+O D+D+ QG C+A
Sbjct: 31 SVLSYTDVQKDSHNLNIPQTPRQHGLT TTPQKLP SHKSPQKQKDSQHQGQHGCLA 90
Query: 72 NGVNAAQHMECEKEAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTDSCDQW 131
NGV AAO+QMECE EK A LS +T Q + D H+NG R+ET T AS T+S D N
Sbjct: 91 NGVAAQSQMECETEEAALS PETDTQTAAASPDHVLNGVRNETTDSASSVTNSHDEN 150
Query: 132 ASDSSYRTPGIGVPLPLEE RGAETETKVO ERENGESPLE LQLOQHHEM KVEHE 185
A DSS RT G LP +E E ++QERENG S L LQQHHE+K +E
Sbjct: 151 ACDSRCRTQOTDLGLPSKEGEPVIEALQERENGSLTEGLNPLQHQHEVKETNE 204

Pedant information for DKFZphtes3_72k15, frame 3

Report for DKFZphtes3_72k15.3

[LENGTH] 188
[MW] 20388.32
[PI] 4.62
[HOMOL] TREMBL:AF038388.1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds. 2e-38
[KW] All Alpha
[KW] SIGNAL PEPTIDE 16
[KW] LOW_COMPLEXITY 12.77 %

SEQ MFSCFLCLISFSSLSNYSYDLKESAVNLMA PRTPGRHGLT TTPQKLLSQ HLPQHQGNDT
SEG .XX
PRD ccc
SEQ DKTGAGTCVANGVNAAQHMECEKEAATLSSDTSIQASEPLLDTHIVNGERDETATAP

PCT/IB00/01496

(No Prosite data available for DKFZphtes3_72k15.3)
(No Pfam data available for DKFZphtes3_72k15.3)

DKF2phtes3_72p16

group: intracellular transport and trafficking

DKF2phtes3_72p16 encodes a novel 796 amino acid protein with very strong similarity to Mus musculus maternal-embryonic 3 (Mem3) gene.

Mem3 was isolated from a partial subtraction library of mouse unfertilized eggs and preimplantation embryos. Its transcript is abundant in the unfertilized egg and also actively transcribed from the newly formed zygotic genome. As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (Pca), proteinase B (Prb), and alkaline phosphatase (ALP).

The new protein can find application in modulation the sorting of proteins into different compartments.

strong similarity to mouse MEM3 and yeast VPS35

Sequenced by DKF1

Locus: /map="16p13.3"

Insert length: 2707 bp

Poly A stretch at pos. 2697, no polyadenylation signal found

```
1 CTACGCGCGG GCGCGGTGCT GCTTGCTGCA GGCTCTGGGG AGTCGCATG
51 CCTACAACAC AGCAGTCCCC TCAGGATGAG CAGGAAAAGC TCTTGGATGA
101 AGCCATACAG GCTGTGAAGG TCCAGTCATT CCAAAATGAG AGATGCGCTGG
151 ACAAAAACAA GCTTATGGAT TCTCTAAAC ATGCTCTAA TATGCTGGT
201 GAACTCCGGA TTCTATGTT ATCCCAAG AGTTACTATG AACTTTATAT
251 GCGCATTTCT GATGAACGTC ACTACTGGA GGTCACCTG ACAGATGAGT
301 TTGCTAAAGG AAGGAAAGTG GCAGATCTCT ACGAATCTGT ACAGATGCT
351 GGAACATTTA TCCCAAGGCT TTACTCTTGT ACACAGTTG GAGTGTATTA
401 TGTCAGTCA TTTCTCTAGT CCAGGAAGGA TATTTGAAA GATTTGTAG
451 AAATGTGCGG TGGTGTGCAA CATCCCTTGA GGGTCTGTT TCTTGAAT
501 TACCTTCTTC AGTGTACCAG AAATATCTTA CCGATGAAG GAGAGCCAAC
551 AGATGAAGAA ACAACTGGTG ACATCAAGTA TTCAATGAT TTCTACTGC
601 TCAACTTTGC AGAATGAAC AAGCTCTGGG TCGAATGCA GCATCAGGGA
651 CATAGCCGAG ATAGAGAAA AAGAGAACGA GAAGACACAG AACTGAGAT
701 TTTAGTGGGA ACAATTTGG TCGGCTCAG TCAGTTGGA GGTGTAAATG
751 TGAACGTTA CAACAGAT TTCTTCACTG GCATATTGGA GCANCTGTA
801 AACTGTAGGG ATGCTTTGGC TCAAGAAAT CTCTAGAGT GTATTATCA
851 GOTTTCCTCT GATGAATTC ACCTCCAGC TTGGAATCCT TTCTCTGGG
901 CCTGTGCTGA GTTACACCAG AATGTAATG TGAAGACAT AATCATGCT
951 TTAATTGATA GATTAGCTT ATTTGCTCAG CGTGAAGAT GACCTGGAAT
1001 CCCAGCGGAT ATTAACCTT TTGATATATT TTCACAGCAG GTGGCTACAG
1051 TGATACAGTC TAGACAGAC ATGCTTTCAG AGGATGTTGT ATCTTTACAA
1101 GTCTCTCTGA TTATCTTGG CATGAATGT TACCCTGATC GTGTGACTA
1151 TGTGTATAAA GTTCTAGAAA CAACAGTGGG GATATTCAAT AAGCTCAACC
1201 TTGAACATAT TGCTACCAAT AGTGCAGTT CAAGGAACCT CACCAGACTT
1251 TTGAATATAC CAGTTGACAC TTACACAAAT ATTTTAACAG TCTTGAATTT
1301 AAAACATTTT CACCCTACTT TTGATGACTT TGACTGAGG TCCAGAAAGA
1351 GCATGAGTTG TTATGTGCTT AGTAATGTTT TGGATTATAA CACAGAAATT
1401 GTCTCTCAAG ACCAGGTGGA TTCCATAATG AATTGTGAT CCACGTTGAT
1451 TCAAGATCAG CCAGATCAAC CTGTAGAAGA CCTGTATCCA GAAGATTTTG
1501 CTATGAGGGA GAGCTTTTGG GGGCTCTTCA TTATCTGCT GGGCTCTGAG
1551 GACCCTGACC AGCAGTACTT GATTTTGAAC ACAGCACGAA AACATTTTGG
1601 AGCTGTGGA AATCAGCGGA TTGCTTTCAC ACTGCCACCT TTGGTATTTG
1651 CAGCTTACCA GCTGGCTTTT CGATATTAAG AGAATTTCAA AGTGGATGAC
1701 AAATGGGAAA AGAATGGGCA GAGCATTTT TCAATTTGCC ACCAGACTAT
1751 CAGTCTTTTG ATCAAGCAG AGCTGGCAGA ATTGCCCTTA AGACTTTTTC
1801 TTCAGGAGC ACTAGCTGCT GGGGAATTTG GTTTTGAAA TCATGAGACA
1851 GTCCGATATG AATTCATCTC CAGGCAATTT TCTCTATGAT AAGATGAAT
1901 CAGCGATTCG AAAGCAGAGC TAGCTGCCAT CACCTTGATC ATTTGGCATT
1951 TTGAAGGAT GAATGCTTTC AGTGAAGAGA ATCATGAACC TCTGAGGACT
2001 CAGTGTGCCC TTGCTGCATC CAAACTTCTA AAGAACCCTG ATCAGGGGCG
2051 AGCTGTGAGC ACCGTGCGAC ATCTCTCTCT GTCTGGGAGA AACAGGACA
2101 AAATGGGGA GGAAGCTTAC GGAGGCAAGA GGGTAATGGA CTGCTTAAAA
2151 AAAGCTCTAA AAATAGCAAA TCAGTGCATG GACCCCTCTC TACAAGTGCA
2201 GCTTTTATTA GAAATTTCTA ACAGATATAT CTATTTTAT GAAAGGAAAA
2251 ATGATGCGGT ACAAATTCAG GTTTTAAAGC AGCTATCCA AAGATTCTGA
2301 GAGACCTCC CGAATCTTGA ATCCAGTGAA GAAACAGAGC AGATTACAA
2351 ACATTTTCAT AACACACTGG AGCATTTGCG CTTCGGGCGG GAATCACCAG
2401 AATCCGAGGG GCCAATTTAT GAAGGTCTCA TCCTTAAAA AGGAAATAGC
2451 TCACCATACT CTTTTCATG TACATCAAT GAGGTTTIA TTACGTAGG
2501 TTTCCCTTCC ATAGATTGTG CCTTTCAGAA ATGCTGAGGT AGGTTTCCCA
```

WO 01/12659

PCT/IB00/01496

2551 TTCTTACCT GTGATGTGTT TTACCCAGCA CCTCCGGACA CTCACCTTCA
2601 GGACCTTAAT AAAATTATTC ACTTGGTAAG TGTTCAAGTC TTTCGTATCA
2651 CCCCAGTAG CATGACTGAT CTGCAATTTA AAATTCCTGT GATCTGTAAA
2701 AAAAAAA

BLAST Results

Entry AC007223 from database EMBL:
Homo sapiens chromosome 16 clone 480g7, WORKING DRAFT SEQUENCE, 38
unordered pieces.
Score = 1081, P = 2.8e-217, identities = 219/221
13 exons

Entry HS015146 from database EMBL:
human STS WI-8848.
Score = 2033, P = 2.9e-87, identities = 425/436

Medline entries

96327632:
Genetic mapping and embryonic expression of a novel, maternally
transcribed gene Men3.
97258867:
Endosome to Golgi retrieval of the vacuolar protein sorting receptor,
Vps10p, requires the function of the
VPS29, VPS30, and VPS35 gene products.
92360909:
Alternative pathways for the sorting of soluble vacuolar proteins in
yeast: a vps35 null mutant missorts and
secretes only a subset of vacuolar hydrolases.
10198044:
Distinct Domains within Vps35p Mediate the Retrieval of Two Different
Cargo Proteins from the Yeast
Prevacuolar/Endosomal Compartment

Peptide information for frame 3

ORF from 48 bp to 2435 bp; peptide length: 796
Category: strong similarity to known protein
Classification: unset

1 NPTTQSPQD EKEKLLDEAI QAVKVSQFM KRCLDENKLM DSIKHASHML
51 GELRTSMLEP KSYELYNAT SOELHYLEVY LTDEFAGRK VADLYELVQY
101 AGNIIPRLYL LITGVVYVK SFPQSRKDL KDLVENCROV QHPLRGLFLR
151 NYLLOCTRM LPEGEPTDE ETTGDISDM DFVLLWFAEN NKLMVVMQHO
201 GSRDREKRE RENOELAILV GTNLVLSOL EGVNVERVSK IYLTGILEQV
251 VNCRDALAQE YLNECIIQVF RDEFHLOTLM PFLRACAEIH QMNVKNII
301 ALIDRIALFA HREDGPGIPA DIKLFDFSQ QWATVIQSRQ DMPSDDVSL
351 QVSLINLAWK CYDRDVQVDV KVLDTTVEIF NKLWLEHAT SSVSKELTR
401 LKLEPDTYN NLTVLKLEH FKLFEYFDY ESRKSHSCYV LSVLDYNTG
451 IVSQDQVDSI MNLVSTLIQD QPDQVEDPD PEDFAEQSL VGRFHLRLS
501 EDPDQVYLIL NTKRHFGAG GNRIRFTLP PLVFAAYQLA FRYKENSQVD
551 DRNKKQKQI FSPNQITSA LKALAELEP LRLFLQALA AGEIGEMNE
601 TVAYEFMSQA FSLYEDEISD SKAQLAAITL IGTTFMKC FSEENHEPLR
651 TQCALAASKL LKPDQGRAV STCAHLFWSG RNTDKNGEEL HGGKRVMECL
701 KKALKIANOC MDPSLQVOLF IEILNRYIYF YEKENDAVTI QVLNQLIQKI
751 REDLPWLESS EETEOINKHF HNTLEHLRLR RESPESEGP1 YDGLIL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphes3_72p16, frame 3

TREMBL:AF024504.3 gene: "A_TH017A05.7": Arabidopsis thaliana BAC
TH017A05., N = 2, Score = 327, P = 1.9e-162

PIR:S56936 vacuolar protein-sorting protein VPS35 - yeast
(Saccharomyces cerevisiae), N = 3, Score = 826, P = 1.5e-116

TREMBL:MH47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds., N = 1, Score = 3376, P
= 0

TREMBL:S42186_1 gene: "VPS35"; product: "Vps35p"; VPS35= vacuolar
protein sorting (Saccharomyces cerevisiae=yeast, Genomic, 3790 nt), N =
3, Score = 813, P = 4.4e-115

>TREMBL:MH47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds.
Length = 754

HSPs:

Score = 3376 (506.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 666/721 (92%), Positives = 682/721 (94%)

Query: 78 EVYLTFEFAKGRVADLYELVQYAGNIIPRLYLITVGVVYKSPQSRKDIKDLVEMC 137
+VYLTFEFAKGR++ADLYELVQY+GHIIIPRLYLITVGVVYKSPQSRKDIKDLVEMC
Sbjct: 34 KVLTFEFAKGERLADLYELVQYSGHIIIPRLYLITVGVVYKSPQSRKDIKDLVEMC 93

Query: 138 RGVOHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLIVRM 197
RGVOHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLIVRM
Sbjct: 94 RGVOHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLIVRM 153

Query: 198 OHQGHSDREKREKREKRELRLVGTNLVLSQLEG-VHVERYKQIVLTGILEQVVNCRDA 256
OHQGHSDREKREKREKRELRLVGTNLV L+ + +QIVLTGILEQVVNCRDA
Sbjct: 154 OHQGHSDREKREKREKRELRLVGTNLVLTLSVRCKCOTLOOTVLTGILEQVVNCRDA 213

Query: 257 LAQEVLMCEIIVQVDFPHLQTLNPFRLACAEHQNVNVMIIIALIDRLALFAHREDGP 316
LAQE KECIIVQVDFPHLQTLNPFRLACAEHQNVNVMIIIALIDRLALFAHRE P
Sbjct: 214 LAQESMECIIVQVDFPHLQTLNPFRLACAEHQNVNVMIIIALIDRLALFAHREDGP 273

Query: 317 GIPADIKLFDIFSQQVATVIQSRQDMPSDDVSLQVSLINLAMEKCPORVDVVDKVLETT 376
GIPA++KLFDIFSQQVATVIQSR+DMPSEDVSLQVSLINLAMEKCPORVDVVDKVLETT
Sbjct: 274 GIPAEIKLFDIFSQQVATVIQSRQDMPSDDVSLQVSLINLAMEKCPORVDVVDKVLETT 333

Query: 377 VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYESR--K 434
VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYES K
Sbjct: 334 VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYESPGK 393

Query: 435 SMSCVVLSNVLDTNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQDFADEQSLVGRF 494
SMSCVVLSNVLDTNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQDFADEQSLVGRF
Sbjct: 394 SMSCVVLSNVLDTNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQDFADEQSLVGRF 453

Query: 495 IHLRSDDPQQOYLINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENSVDXWE 554
IHLRS+DPQQOYLINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENS K
Sbjct: 454 IHLRSDDPQQOYLINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENSXWTSKG 513

Query: 555 KKCQKIFSFHQITISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEYMSQAFSLY 614
+ ++ F HQTISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEYMSQAFSLY
Sbjct: 514 RMRRYFHLHQITISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEYMSQAFSLY 573

Query: 615 EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRTQCALAASKLLKPKQGRAVSTCA 674
EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRT+CALAASKLLKPKD C
Sbjct: 574 EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRTQCALAASKLLKPKQAEERHMT 633

Query: 675 HLPWSGRNTDNGEELHGGKRVNCECLKALKIANQCHDPSLQVLFIEILNRYIFYEKE 734
L WSGRNTDNGEELHGGKRVNCECLKALKIANQCHDPSLQVLFIEILNRYIFYEKE
Sbjct: 634 SL WSGRNTDNGEELHGGKRVNCECLKALKIANQCHDPSLQVLFIEILNRYIFYEKE 692

Query: 735 NDAVTIQVNLQIQIRDELPHLESSEETQINKHFNHTLEHLRRSPESSEGPIVEGL 794
NDAVTIQVNLQIQIRDELPHLESSEETQINKHFNHTLEHLRRSPESSEGPIVEGL
Sbjct: 693 NDAVTIQVNLQIQIRDELPHLESSEETQINKHFNHTLEHLRRSPESSEGPIVEGL 752

Query: 795 IL 796
IL
Sbjct: 753 IL 754

Pedant information for DKF2phtes3_72p16, frame 3

Report for DKF2phtes3_72p16.3

(LENGTH) 796

930

PCT/IB00/01496

```

SEQ      RESPESEGPIYEGLIL
SEG      .....
PRD      hccccccccceccc
MEM      .....

```

(No Prosite data available for DKFZphtes3_72p16.3)
(No Pfam data available for DKFZphtes3_72p16.3)

DNFzphes3_7b22

group: cell structure and motility

DNFzphes3_7b22 encodes a novel 443 amino acid protein with weak similarity to paramyosins.

The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

similarity to paramyosins

complete cDNA, complete cds, few EST hits

Sequenced by BNF2

Locus: /map="3"

Insert length: 2291 bp
Poly A stretch at pos. 2241, polyadenylation signal at pos. 2213

```
1 GGAAGAAAGG CTAGCGGGCG TTGGCCGTAT GTGGGTGCTT TGAGGCAGTT
51 TTTCAGTTCT TTCATTACC AAGTGACAT GCACCTACTA GGTGCCAGGT
101 GTTTTGACGT ACATACAACC CTTCGAAA CTCTTCAGTG TAGTCTCTCG
151 TATGAAAAGT TTCCAGCCAA GAATGCCAC TGCACCTGAG ATAAGGGGGA
201 TCCTGGCCAT TAAGGAACCC TTGGCTTCGA AACTGAGCCG TGAGGAACCTA
251 TACAAATGG GAAATGGGA CAATCCGAG TGCTCATGA CACTAAGAAG
301 TAAATTTACG AACTCACTGA GCTGGAAGTC ATTCACGGG AATTGAATAG
351 GTAAGTGCAC TTTTGTGAGA TTATAAATAT ACCACGGAG GTAACGAAGC
401 TACAGAAGAA TGAAGAAGA CAGCTCGAA GACTCAAAAC TTCTCCAAA
451 AGTTTGCAAT TCTGAGTGA CGGTGTCAGT GACAGCGGAA CCACCTAGTA
501 CGGTAGAAGA AGAAGGAATA CCTAAAGAAA CAGACATAGA AATCATCCCA
551 GAAATCCCGG AAACCTCTAG GCACTGTCC CTTCAGATG TGCTGAGGAT
601 CTGGGCACT CTGGAGGACA CACAGACCA GCTCTCTATT CTGAACATGA
651 TCATGCCCGT TCAGTACGAA GGGAGACAGA GCATCTGCTT GAAAGCAGA
701 GAATGAATC TAGAAGGAAC GAATCTAGC AAATTCCAA TGGCTCAAC
751 AATCAGAAA ATACCCAGTC GCTTAATAC TGAGGAAGGA CCCAATTGC
801 CAGAAATCAG ACACAGAGGC CGTTGCTGT TGAGTTTAA CAJAATGAG
851 GATCTGTCT TCAGAAAACC TACAGGCGAG ACCATCATGA CTACGGAGAC
901 ACTGAAGAAA ATTCAGATTG ATAGGCAGTT TTTACGGAT GTGATTGAG
951 ATACCATTA GAGTTGCAA GATTGCGCA CTACAGAGG TCTCTGCAA
1001 GCTTTGAGA AAGAGAGGA AAKCAAAATG CATTTATG ACATCATTCG
1051 CAGGAGGAA AAGGAAGAA AACAGATAAT ATCACTCAA AAACAGCTAA
1101 TTAATGTCAA AAAGGAATGG CAATTTGAAG TCCAGATCA GAATGATAT
1151 ATGCTAACC TCAGAGACA ACTGCAAGAG ATGAAGGAA ATCCACTT
1201 GGAGATCGC TACATGAAA CCAATACCGA GCTGCAGATT GCCAGACCC
1251 AGAAGAAAGT TAACAGAACA GAGGAACCTT TGGTGAAGA GATTGAGAAA
1301 CTCAGATGA AAACCGAAGA AGAGGCCCGG ACTCATACAG AGATTGAAT
1351 GTTCTTAGA AAGAGGACG AGAAGCTTGA GACAGGCTG GACTCTGCA
1401 TGGAGAAATA CGATAAGGAC ACAGAAATGA AACAGATGA ACTAATGCT
1451 CTCAGAGCCA CAAGGCCAG TGACTTAGCA CACCTTCAAG ACCTGGCAAA
1501 GATGATAGA GATATGAAK AGGTATCAT TGAAGATGCT ATAGAAAGG
1551 AGAGGACCA GAGAGAGTA AAKAGGATC TCTTGAATT AAAGAGCCTT
1601 ATAAAGCTCC AGGCTGGTG GAGAGGACT ATGATACGA GAGAAATTGG
1651 TGTTTCAAG ATGCTTAAG ACAAGTTGA TAGCAGGAT TCAAAAGGCA
1701 AAGTTAAGG CAGGATAG AGGAGAGGCA AGAAGAGTG ACCAGTTCT
1751 CTTTGTGTT TTCTGCTGT ATTCTGGAG TGGAAGGAC TTGAGAGTT
1801 AAGAAACACC TGTACCTCA AAGATGACT ATCTACAGT TTTTCTCTAT
1851 TGAGACTTTC CCAGGGAAG CTGATTTCAC TTGCTGTIT AATTTCACCT
1901 TGCTGTITG GTGGTTTTT AACCTCTAT TTAGATTAC ACATTCGCT
1951 TAGGGCTTCC TCATACCTTG CTGGGAAGAA GTTCTAGTA GTCTGTGAA
2001 GATTCTTCT TCTGCTCTT TCTCAGAGA ACAAGGAGT TCAGTGCTT
2051 AGCTACACTC ACCATTGAA ACTGAGTAA TTCTGTATAT GTCAATTTT
2101 CATTTTACCC AATTGTCTG TACTGAAAA ACCTTATGA GCAAAATAT
2151 TCAGTAGGAA TTACAATAG ATGTTATTAG CTGTCAGCA TAATATATAC
2201 ACAGCAAGT TTAATAAAT GTTGGTCTT GCTGCTTTT TAAAAAATAA
2251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry G36731 from database ENBL:
SHGC-52923 Human Homo sapiens STS cDNA.

WO 01/12659

PCT/IB00/01496

Score = 2262, P = 1.3e-97, identities = 462/468

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 410 bp to 1738 bp: peptide length: 443
Category: similarity to known protein

```
1 MEEDSLEDN LPKVVHSEM TVSVTGEPPS TVEEGIPKE TOIGIIPFIP
51 ETEPLESLPD VLRISAVLED TTDQLSILMY IMPVOYEGRO SICVKSADNM
101 LEGTNLDKLP MASTITKI9S PLITEEGPWL PEIRHGRITA VEINRMODLV
151 FKPPFQTQIM TYTELAKIQI DROFFSOVIA DTIKELQOSA TVVSLQALS
201 KERENKHIFY DIAREEKGR KQIISLQKOL INVKKEMQTE VQSQNEYIAN
251 LKDLQENKA KSNLENRYNK THTELQIAQT QNCKNTEEL LVEETELRM
301 KTEEEARTHT ELENFLAREQ QLEEKLEFW HERYDKDTEM KQNELNALRA
351 TRASDLAWLQ DUAHMIREY OVITEERIEK ESKKVKQOO LLEKSVIKL
401 QAWWRGTMR REIGGFMPK DKVDSKDEK GKKGKRRG KKK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRFZphtes3_7b22, frame 2

SWISSPROT:MYSP_BRUMA PARAMYOSIN., N = 1, Score = 158, P = 5.8e-08
PIR:A44972 paramyosin - nematode (Dirofilaria immitis) (fragment), N = 1, Score = 157, P = 7.1e-08
SWISSPROT:MYSP_OMCVO PARAMYOSIN., N = 1, Score = 157, P = 7.4e-08
PIR:S52537 emm L 15 protein - Streptococcus pyogenes, N = 1, Score = 151, P = 8.6e-08

>SWISSPROT:MYSP_BRUMA PARAMYOSIN.
Length = 880

HSPs:

Score = 158 (23.7 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 66/259 (25%), Positives = 125/259 (48%)

Query: 142 EFNKQQLVFKKPTQIMTTETLKKI9IDROFFSDVIADTIKELQDSATYNSILQALS 201
+ K + L R R T E K++ + +D +A + LQ A N LL+ +
Sbjct: 169 QLKKEHLEAKAERFEAQTVELSNKVEDLNHRVND-LAQQRQLQ--AENDLLKEIND 225
Query: 202 ER---ENKMHF-YDIAREEKGRKQIISLQKOLINVKKEMQFEVOSONEYIANLKQQLQE 257
++ +N H Y + +E+ R+++ +++ ++ + +VQ + + + D+ E
Sbjct: 226 QKVQLDNLQHVKYOLAQOLEEARRRLEDAERERSQLQAQLH-QVQLEDSVPTALDE--E 282
Query: 258 MKAKSNLENRYNKTHTELQIAQTQKCKNTEELLVEEIEKLMKT-EZEARTHTIEHFL 316
A++ E++ NTE I Q + K + L EE+E LR K +++A +IE+ L
Sbjct: 283 SAARAEEHKLALANTE--ITQWKSFTDAEVALHHEEVEDLRKMLQKQAEYEEQIEIML 340
Query: 317 RKEQO--KLEERLEFWHEKYDKDTEMKQNELNALKATKASDLANLQDLAKMIREYEQV 374
+K Q K + RL+ +E D E QN + L+ K + L K + E + I
Sbjct: 341 QKISOLEKAKSRLOSEVEVLVDLEKAQNTIILERAK-----EQLEKTVNELKVRID 393
Query: 375 EDRIKERSKRVKQOLLEKSVIKL 400
E +E E +++ + L EL+ + L
Sbjct: 394 ELTVELEAAQREARAAELQKLKHL 419

Score = 118 (17.7 bits), Expect = 1.3e-03, P = 1.3e-03
Identities = 54/231 (23%), Positives = 108/231 (46%)

Query: 181 DTIKELQDSATYNSLIQ---ALSKERENKHIFYDIAREEKGR-RKQIISLQKOLINVKK 235
D +KE+ D Lw L+++ E + RE + Q+ +Q +L +V+
Sbjct: 218 DLLKEIHDKVQLDNLQHVKYOLAQOLEEARRRLEDAERERSQLQAQLHVOLEDSVRT 277

Query: 236 EVQPE--VQSQNEY-IAHLKQDLQENKAKSLNRYNKTNT-LOIAQTKCKNTEELL 291
 E + + + E + A + + + K + K + E E L + QK + E + +
 Sbjct: 278 ALDESAARAAEAHKLALANTEITOMKSKFDAEVALHHEEVEDLRKKMLQKQAEYEEQIE 337

Query: 292 VEEIEKLKMTTEEARNTTEIMF---LKEQOKLE--ERLEFNMKEYDKOTENKQELM 346
 + + + K + + + R + E + E + L K Q + ER + + EK + + + + EL
 Sbjct: 338 IM-LQKISQLEKAKSLQSEVEVLVDLEKAQNTIATLERAKOLEKTVNMLKVRIDELT 396

Query: 347 A-LKATKASDLAHLQDLAKMIREYQVILIEDRIEKSKKKVQDLLEKSVI 398
 L + A + A L + L K + YE + E + R KK + + DL E K +
 Sbjct: 397 VELAAQREARAAALAELOKLNLYEKAV-EQKEALARENKILQDDLHEAKAL 448

Score = 107 (16.1 bits), Expect = 2.1e-02, P = 2.1e-02
 Identities = 49/279 (17%), Positives = 124/279 (44%)

Query: 123 IYEGPHLFEIRHGRFAV-EFNHMOQLVKKPTROTINTTETLKKIQIDRQFFSDVIAD 181
 I E L + R A + E K + + L K + + + E KK + Q D + + AD
 Sbjct: 392 IDLTVELEAAQREARAAALAELOKLNLYEKAVEQEALAREN-KKQDDLHEAKALAD 450

Query: 182 TIKELQDSATYNSLLQALSKEKHNKHFTDIAREKGRKQ--IISLOKOLNVKKEWQF 239
 + + L + N + L + E + + + R + R Q + LQ + I + + + Q
 Sbjct: 451 ANRKLHLDLENARLAGEIRELQALKSEAAARRDAENRAQALAELOQLRIEMHRLQE 510

Query: 240 EVQSQNEYIANLKQDLQENKAKSLNRYNKTNT-LOIAQTKCKNTE-ELLVEIEKL 298
 + + N + + + A L + + E + + + E E + V + + +
 Sbjct: 511 KEEMEALRKNMFEIDRLTAA--LADAKARKAEISRLKKYQAEIAELHTVDHLRA 568

Query: 299 RKTEEEARNTTEIMFLEKQKLEERLEFNMKEYDKOTENKQELMALKATKASDLAHL 358
 + + + + + E L + + + + + + + + + + Q + + + AL A + + +
 Sbjct: 569 NIEAKYIKKQSEQLK:LOASLEDTQKQQLDQY----ALAGRKVSALSA-ELEECKV 623

Query: 359 LQDLAKMIREYQVILIEDRIEKSKKKVQDLLEKSVIRLQ 401
 D A R + + + E + + + V + L + K + + + +
 Sbjct: 624 ALDNATIRAKQAEIOLEANGRITOLVSVMNLTATIKHLETE 666

Pedant information for DKF2phtes3_7b22, frame 2

Report for DKF2phtes3_7b22.2

[LENGTH] 443
 [MW] 51917.95
 [PI] 6.18
 [HOMOL] PIR:S28589 trichohyalin - rabbit 2e-08
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-07
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 1e-07
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M. jennaschii, MJ1322] 5e-06
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 3e-05
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR055w] 6e-05
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR055w] 6e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-04
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL201c] 5e-04
 [EC] 3.6.1.32 Myosin ATPase 3e-08
 [PIRKM] phosphotransferase 6e-06
 [PIRKM] citrulline 8e-06
 [PIRKM] tandem repeat 1e-07
 [PIRKM] heart 6e-06
 [PIRKM] polymorphism 4e-06
 [PIRKM] serine/threonine-specific protein kinase 6e-06
 [PIRKM] DNA binding 8e-08

[PIRKM] muscle contraction 1e-07
[PIRKM] actin binding 3e-08
[PIRKM] ATP 3e-08
[PIRKM] thick filament 1e-07
[PIRKM] phosphoprotein 3e-08
[PIRKM] glycoprotein 4e-06
[PIRKM] skeletal muscle 1e-07
[PIRKM] calcium binding 8e-06
[PIRKM] alternative splicing 3e-08
[PIRKM] coiled coil 3e-08
[PIRKM] P-loop 3e-08
[PIRKM] heptad repeat 4e-06
[PIRKM] methylated amino acid 3e-08
[PIRKM] basement membrane 4e-06
[PIRKM] cardiac muscle 6e-06
[PIRKM] extracellular matrix 4e-06
[PIRKM] hydrolase 3e-08
[PIRKM] membrane protein 4e-06
[PIRKM] EF hand 8e-06
[PIRKM] cytoskeleton 8e-06
[PIRKM] hair 8e-06
[SUFFAM] myosin heavy chain 3e-08
[SUFFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-06
[SUFFAM] calmodulin repeat homology 8e-06
[SUFFAM] myosin motor domain homology 3e-08
[SUFFAM] trichohyalin 8e-06
[SUFFAM] protein kinase homology 6e-06
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 4
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All Alpha
[KW] LOW_COMPLEXITY 10.61 %

SEQ MEEDSLEDGNLPKVVNSDHTVSVTGEPPSTVEEGIPKETDIEIIPETLEPLSLPD
SEGXXXXXXXXXXXXXXXXXXXXX
PRD CC
SEQ VLRIASVLEDTTQGLSILWYIMPVQYEGRSICVKSREMNLEGTNLDKLPMASTITIKIPS
SEG chhh
PRD
SEQ PLITEEGPHLEIRHGRFAVEFNMQDLVFKKPTQIMTTETLKKIQIDRQFFSDVIA
SEG
PRD hhh
SEQ DTIKELQDSATYNSLLQALSKERENHMFYDIAREEKGKRIISLQQLINVKKEWQFE
SEG
PRD hhh
SEQ VQSQNEYIANLKQQLQEMKASNLNRYMKTNTLQIAQTQKCNRTTEELLVEEIEKLRM
SEG
PRD hhh
SEQ KTEELARTHTIEIMFLRKEQKLEERLEPWMEYKQDTEMKQELNALKATKASDLARLQ
SEG
PRD hhh
SEQ DLAKMIREYEQVIEDRIEKERSKKKVKQLLEKSVIKLQAWRGTHIRREIGGFMMFK
SEG
PRD hhh
SEQ DKVDSKDSKGGKGGKDKKGGKKK
SEG
PRD ccccccccccccccccccccc

Prosite for DKFzptes3_7b22.2

PS00001	285->289	ASN_GLYCOSYLATION	PD0000001
PS00004	152->156	CAMP_PHOSPHO_SITE	PD0000004
PS00005	164->167	PKC_PHOSPHO_SITE	PD0000005
PS00005	182->185	PKC_PHOSPHO_SITE	PD0000005
PS00005	280->283	PKC_PHOSPHO_SITE	PD0000005
PS00005	383->386	PKC_PHOSPHO_SITE	PD0000005
PS00006	5->9	CK2_PHOSPHO_SITE	PD0000006
PS00006	30->34	CK2_PHOSPHO_SITE	PD0000006

WO 01/12659

PCT/IB00/01496

PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	182->186	CK2_PHOSPHO_SITE	PDOC00006
PS00006	243->247	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	310->314	CK2_PHOSPHO_SITE	PDOC00006
PS00007	261->269	TYR_PHOSPHO_SITE	PDOC00007
PS00007	184->193	TYR_PHOSPHO_SITE	PDOC00007
PS00009	218->222	AMIDATION	PDOC00009
PS00009	439->443	AMIDATION	PDOC00009

(No Pfam data available for OMFphtes3_7b22.2)

DKFZphtes3_7d17

group: testes derived

DKFZphtes3_7d17 encodes a novel 633 amino acid protein with weak similarity to human KIAA0454.

Pfam predicts a TNFR/MGFR cysteine-rich region.

No informative BLAST results: No predictive prosite or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0454

complete cDNA, complete cds, EST hits

Sequenced by BHFZ

Locus: unknown

Insert length: 3608 bp

Poly A stretch at pos. 3587, polyadenylation signal at pos. 3570

```
1 GGGAGTTAC GCGGAAGTCC ACCCAGCGTT TCTCAGGCAA TCTGAAGSCA
51 AATCCTGTTT AGCCCAAGGC GAAGGTTCTT GGTGACCCAG GCTCTCACC
101 GCCAATTGTC CTTTCCCGTC CTCTGAGGG TATCTGGAGC TTCAGTGGTG
151 TGTGCTTTT GCTCCACAC TGGGGAATCC ACTGATGCC ACTGTCCAGG
201 GCTTCCAGTG GACTCTCCGA GGCCCTGATG TAGAACTTC CCCATTCGGT
251 GCACCAAGAG CAGCCTCACA TGGTGTGGGG CGACATCAAG AGCTGCGAGA
301 TCCACACAGT CCTGCGCCCA CCTCTTCTGG CACAAGAGTC AGCATGGTGG
351 TATCTGCCGG CCTTGGTCTG GTTGAGAGAG CAGAGATGAA CATCTTAGAA
401 ATCAACAGAA AATCGGCGCC CCAGCTGGCA GAGAACAAAC AGCAGTTTCA
451 AAACCTCAAA CAGAAATGTC TTGTAACCTA AGTGCCCTAC TTCTGGGCCA
501 ACCCGCAAAA TATTTAGAG TATGAGAGCT GCAGAGAGCT CATTAATCT
551 ATGCTGAGGG ATGAGCGGCT GCTCAGAGAA GAGAAGCTTG CAGAGGAGCT
601 CGGCAAGCT GAGGAGCTCA GGCAATATAA AGTCTGTGTT CACTCTCAGG
651 AACGAGAGCT GACCCAGTTA AGGGAAGAT TACAGGAGG CAGAGATGCC
701 TCCCGCTCAT TGAATCAGCA TTCTCAGGCT CTCTCTACTC CGATGAGGCC
751 GGACAACTCC CAGGGAACGG ACCTCCGAGA ACAGCTGGCT GAGGGATGTA
801 GGCTGGCACA GCACCTCTGC CAAAAGCTCA GCCCAGAAA TGTATGAGAT
851 GAGGATCAAG ATGTTAAGAT TGAGGAGGCT GAGAAAGTAC AGGATTTATA
901 TGCCCGCAGG GAGGTGCGAG AGGCTGAAGA AAAGGAAGTC CCTGAGGACT
951 CACTGGAGGA GTGTGCCATC ACTTGTTCAA ATAGCCACCA CCCTGTGTAG
1001 TCCACACAGG CTTACGGGAA CACCAGAATC ACATTTGAGG AAGACCAAGT
1051 CGACTCAACT CCAATTGACT CATGCTCTCA TGATGAATGG TTGCACTGCT
1101 TATGCAATT CCGCAGAAAT GAAAGTGATC ATGAGCAAGA GGAGAAAAAA
1151 GGGCCAGTGT CTCCAGGAA TCTGAGGAG TCTGAAGAG AGGAAGCCCC
1201 CCAGAGTCC TGGGATGAAG GTGATTGAG CTCTCAAT CTCTTGACA
1251 TGTCTGCTC ATCCAGTCT GACAGAGCA CCTTTCAC TCATGAGGAA
1301 CAGCAAGTCC GCTTGGCTCT TGACATAGGC AGACATGGT GTGATCAAGT
1351 GAAAAAGGAG GACCAAGAGG CCACAGTCC CAGGCTCAGC AGGAGCTGCT
1401 TGGATCAGAA AGAGCTGAAA GTCTTCAGG ACTCAGTCA TAGATTTAT
1451 TCAACTCCTT TTGAGTACCT GGAAGTCCCT GACTTATGCC AGCCCTACAG
1501 AAGTGACTTT TACTCATTCG AGGAACACA CTTTGGCTTG GCTCTTGACT
1551 TGGCAGAAAT GAAAAAGAC CAGAGAGAG AGAGAGCCA AGGCCACCA
1601 TGCCCAAGG TCAGCAGAGA GCTGCCGAG GTAGTAGAGC CTGAGGACTT
1651 GCAGGACTCA CTGGATAGAT GGTATTGAC TCCTTTCACT TATCCAGAAC
1701 TGCTGTATTC ATGCCAGCCC TACGGAGATT GCTTTTACT ATYGAGGAA
1751 GAACAGTTG GCTTTTCTT TGACGTGAT GAATTTGAAA AGTACCAAGA
1801 AGGGGAAGAA GATCAAAAGC CACCATGCC CAGGCTCAAC CAGGTGTGTA
1851 TGGAGCAGAA AGAGCCTGAA GTCTTCAGG ACTCAGTGA TAGATGTTAT
1901 TGGACTACTT CAAGTTACTT TCAATCAT CTCTCATCC AGCAGTACG
1951 AAGTGCTTT TACTCATTTT AGGAACAGGA CGTCAGCTTG CCCTTGACG
2001 TGGACAATAG GTTTTTACT TTGACAGTGA TAAGGCACCA CCTGGCCCTC
2051 CAGATGGGAG TCATATTCCC ACACTAAGCA GGCCTTACTA AGCTAGAGA
2101 TGTCTTCTC CAGGCGAGGA CCTATAGGCA CATCTAGGTT TGATGAAAC
2151 TGTAGTCCC TTGGAAGCC CAGTCATAG ATGGGAAAGT GGCATGGCT
2201 CTATTCTCT TCTCAGACCA TGCCAGTGGC CACTGTGCT CAGTCTGAAG
2251 ACCTTGGACC CAGTTAGGT GTGACAGTT CACACACCTA TGTAGACAT
2301 GCGGGAGTGT ATCTCCAGA CATCTAAT TTGAACAGAT ATCTCTGGGT
2351 AGCTACAAAG TTCTCAGGG GTTTCATTT GCAGCATGT CTCTGAGCTT
2401 CTATACCTGC TCAAGGTCAG TGTCACTTT GTGTTAGCT CATCCAAAGG
2451 TGTATACCTG GTTCTATGA ACTAACCC ATTCTGTGA TCTCTAGGT
2501 TGCTTTGTT TACTGATCC ATCTGTACA CAGGAGGAT CTTGGCTGA
2551 GGATTGTATT TCAGAACCA TGAAGTCTCT TGACAGTTGT TAACCCACTA
2601 GGCTCCTTTG AGTAGAGAAG CCAATGCTCT TCAGCTTCCA ATTGATATCA
2651 ATACTTAGGA AGACACAGC TAGAGGACA ACAGCATTT GAGGCGCTTA
```

```
2701 GTCTGTGCTCC TTTCATTCCT ATCTGTGTAAG GAACAGGAGT CAGGAGCCGC
2751 TGGCAGAGAGA CAGCATGTCTA CTTGGGACTC TGCCAGTCCA GAATATGAAAC
2801 AATGCCATGT TCTTGCAGAA AATGCTTAGC CTGAGTTTCA TAGAGGTAA
2851 TCACCAGACA ACTGCAGAAAT GTAGAACACT GAGCAGGACA ACTGACCTGT
2901 CTCCTTTACA CAGTCACAGT CACACAGAAAT CACACACAA AAAGGAGGAG
2951 AGATATTTTG GTTTCAGAG AGTAAATGA TAATGTAGCT ACATTTCTTT
3001 AGTATTTTTC AGCCCAAAAT ATTCTCTCAT CTTTGTGTG TTGTCAATTGA
3051 TTTTGTGTGAC ATGGACTTGT TTGTAGAGA CAGTCAGCT GTCTGGGTCA
3101 ATGGTCTACA TTCTGAAGTT GTCTGAAAT GTCTTCATCA TTAATTCAG
3151 CCTAACCTTT TCATCAGAAA CAGTCAGAG TCGATACTCT GAGTTTCCAA
3201 CCTCAGCCCA TCTGTGGCCA GAGAGGTCT AGTTTGTCCA TCAGATTAT
3251 CATGATATCA GGACTGGTTA CTTGTTAAG GAGGGGTCTA GGAGATCTGT
3301 CCGTTTACA GACACTTAC TTATGATGA GTATTGGGA GAGTGGTTT
3351 TCAAGGTAGA AATGTCTGT ATTCCAGTGA TCATCCTCTA AAGTCTTAT
3401 CATTATTAA TCATCCCTGC CTGTGTCTAT TATTATATC ATATCTCTAC
3451 GCTGGAATTT TGTGCTCTCA ATGTTTACTG TGCTTTTGT TTGTGTAGTG
3501 TGTGTGTGTT AAAAAAAN ATCTCTGCTG GAGTTTAA TTTTGTCCA
3551 AAGTATTTT AATCTATACA ATTAATAACT TTGCTTATC AAAAAAANA
3601 AAAAAANA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 176 bp to 2074 bp; peptide length: 633
Category: similarity to known protein

```
1 MPLTPVQGF QNTLRGPDVE TSPFCAPRAA SHGVGRHQEL RDPTVPGPTS
51 SATNVMVMVS AGPWSGEKAE HMLEINKKS RPLAENKQO FRLKQKCLV
101 TVVAYFLANR QNVVDYEDCK DLIKSHLDE ALTEELKAS ELGQAEELRQ
151 TVVLVNSQER ELTQLREKIQ EGRDASRLN QHLQALLTPD EPDMSQGRDL
201 REQLAEGCRL AQHLVOKLSP ENDDDEDEDV KVEAEKVOE LYAPREVOKA
251 EEKEVPEDSL EECATYCSNS RHPCESNQPY QNTRITFEED QVSTLIDSS
301 SHQWDLQAVC ITPENSONE EEEKQVPSF RHLQSEEEZ APQSDWQGD
351 WTLSPFPMMS ASYQSDRSTF NSVEEDQVGL ALDIGRHMCQ QVKKEDQCAT
401 SPRLRELLD EKEPEVLQDS LDRFYSTPFE VLELPOLCOP YRSDPYSLQK
451 QHLGLALDLD RKKHQEKEE DQCPKCPALS RELPEVPEFE QLOSLDMY
501 STFPFPELPS DSCOPYGSCF YSLEEHVGF SLOWDEIKY QEGEEDQKPP
551 CPRLNEVLME ADEPEVLQDS LDRCYSTTST YFOLHASFOQ YRSAPYSFEE
601 QOVSLALDVO HRFPTLTIVR HHILAFQMGV1 FPH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7d17, frame 2

PIR:T00069 hypothetical protein KIAA0454 - human (fragment), N = 1,
Score = 199, P = 1e-11

PIR:A45592 liver stage antigen LSA-1 - Plasmodium falciparum, N = 1,
Score = 158, P = 2.7e-07

>PIR:T00069 hypothetical protein KIAA0454 - human (fragment)
Length = 1,882

HSPs:

Score = 199 (29.9 bits), Expect = 1.0e-11, P = 1.0e-11
Identities = 74/261 (28%), Positives = 122/261 (46%)

Query: 117 EDCKDLIKSHLDRERLLT---EEKLAELGQAEELRQYKVLVHSERELTLQREKIQEG 172
+D + LI+ + + E L EEKLAEL A +Y L+ Q REL+ LR+K+EG
Sbjct: 964 KDLESLIQRVSQLAQLPKNGLEELKLELRSASWPKGYDSLIQQAELSTYLRKIRREG 1023

Query: 173 RDASRLNQH-----LQALLTPDEPDSQGRDLREQLAEGCRLAQHLVQKLSFENODD 225
R + +H + LL + D G+ REQLA+G +L + L KLS +
Sbjct: 1024 RGICYLITRHAKDTVKSFDLLKSHNDIYLGOSFREQLAQGSLTERLTSKLSKDKHS 1083

Query: 226 EDEDVKVEAEKVOLYAPREVQAEAE-EVPEOSLECAITCSNHHPCESHQPTGHTA 284
E + +E L RE+Q+ E+ EV + L+ +T 3+SH +S++ +T
Sbjct: 1084 EKDGAGLEPLA----LRLSRELQEKERVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1139

Query: 285 ITFEEDQV--DSTLDSHSDMDLDAVCIIPENESDHEQEEKGPVSPNQLQSEEEAP 342
+E + D + + +H E A P + +S + S + A
Sbjct: 1140 FLSDLEACSDMDIVSEYTHYEKKAS---PSHSDSHSHSRAVLSKPSSTASQGA 1196

Query: 343 QESWDEGOWTSLIPPDMSASYOSRSTFH 371
ES + +L P + S FH
Sbjct: 1197 AES-WSNPISLPTPONTPEANQANSGFH 1224

Score = 89 (13.4 bits), Expect = 1.1e-01, P = 1.0e-01
Identities = 35/89 (39%), Positives = 44/89 (49%)

Query: 464 KQOEKEEDQG---PCFRLSRELPEVVEP-EDLOSLDRWYSTPSPYFELPQSCQ-PYGS 518
KD + E+DO P RLSREL E + E LQ LD TP S L DS + P +
Sbjct: 1079 KDHSEKDQAGLEPLALRLSRELQEKERVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1138

Query: 519 CFYSLKEHVGSFLOVDEIERVDEGEEDQRP 550
P S E F D+D + + EE + P
Sbjct: 1139 SFLSDELEACS---DMDIVSEYTHYEKKASP 1167

Score = 73 (11.0 bits), Expect = 4.8e+00, P = 9.9e-01
Identities = 31/88 (35%), Positives = 40/88 (45%)

Query: 390 DOVKEDDEATSP---RLSRELLD-EKEPEVLQDSLDRFYSTPFEYLELPOLCO-PYRS 444
D + +DO P RLSREL + EK EVLQ LD TP L D + P +
Sbjct: 1080 DHKSEKDQAGLEPLALRLSRELQEKERVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1139

Query: 445 FYSLQKHLGLALDLMKDKQEEEDQGP 475
F S L D+D + + EE + P
Sbjct: 1140 FLSDLEACSDMDIVSEYTHYEKKASP 1167

Score = 68 (10.2 bits), Expect = 1.1e-01, P = 1.0e-01
Identities = 36/156 (23%), Positives = 68/156 (43%)

Query: 31 SHGVGRHOELRDPV---PCPTSSATNVSMVVSAGPWS-----GEKAMWILEINKK 79
S G +HQE + TV P P S + V A G++ ++ +
Sbjct: 684 SPCKKHQOEGHVTVPFPAPQSLDGLATTVDHQLDMSQPRDGPQSAFSLPGSTQH 743

Query: 80 SRPOLAENKQOFNLKQKCLVTOVAYFL-ANRQNYDYE-DCKDLKSLRDERLLTEEK 137
R OL++ Q++++L++K L+++ F AN Y + L+K + ++ ++
Sbjct: 744 LRSQLSQCKQRYQDLQERLLSEATVFAQAELEKRYVNLTGESLVKQDSRQIQVQLQDL 803

Query: 138 LAELGQAEELQRYK-LVHSQERELTOLREK-LQEG 172
E G++E + + + E L+E L EG
Sbjct: 804 GYETCGRSEKAEETTSPECEHNSLKENLVMS 839

Score = 65 (9.8 bits), Expect = 2.2e-01, P = 2.0e-01
Identities = 23/96 (23%), Positives = 52/96 (54%)

Query: 123 IKSLRDERLLTEEKLAELGQAE-----LRQYKLVHSQERELTOLREKLQEGRDASRS 178
++ + D+ + E + E+ EE LRQ ++ V ++ +L +LR+ L ++ +
Sbjct: 5 LRQRHOKVALEKATCETYSALSEKEKELPOLRLAVRERDHLERDVL-----SNEA 60

Query: 179 LNHQLQALLTPDEPDSQGRDLREQLAEGCRLAQHLVQKLS 218
Q +++LL +G ++ EQL+ C+ Q L +++
Sbjct: 61 THQSMESLL-----RAKLEV-EQLSTTCQLQMLKEEN 93

Score = 61 (9.2 bits), Expect = 5.5e-01, P = 4.2e-01
Identities = 27/95 (28%), Positives = 47/95 (49%)

Query: 134 TEER-LARELGQAEELQRYK---KVLVHSQERELTOLREKLQEGRDASRSLNHQLQALL 188
+E K L +LG+ EE R Y +LV + + + L+ +LQ +L +L
Sbjct: 855 SEARKPLENGKQEEFRVYKSEMILV--LRKDIKLQAKLQANKVIOHLSRVSLV 912

Query: 189 PDEPDSQGRDLREQLAEGCRLAQHLVQKLSFENODDDE 228
+ +S R R+ A G ++ SP + DEED
Sbjct: 913 TSDYSSSLERP-RKLAAVGT-----LEGSSPHSVPEDE 945

Score = 57 (8.6 bits), Expect = 1.4e+00, P = 7.5e-01
Identities = 26/92 (28%), Positives = 47/92 (51%)

Query: 127 LRDERLLTEKLAELGQAEEL---RQYKLVHSQERELTOLREKLQEGRDASRSLNHQL 183
L E LL ER+A Q +E+ R+ +L+ + L R +L S A R L L
Sbjct: 358 LTOEVLLLRKVASVESQOEISGRROQLLNLLEG--LVDESRINELQAEKQVYSSL 415

Report for DKF2phtes3 7d17.2

[illegible]

Prosites for DKFZphtes3_7d17.2

PS00001	54-558	ASN GLYCOSYLATION	POC000001
PS00002	315-319	ASN GLYCOSYLATION	POC000001
PS00003	13-1316	PKC PHOSPHO_SITE	POC000003
PS00004	322-323	PKC PHOSPHO_SITE	POC000003
PS00005	365-368	PKC PHOSPHO_SITE	POC000003
PS00005	401-404	PKC PHOSPHO_SITE	POC000003
PS00006	382-383	CK2 PHOSPHO_SITE	POC000006
PS00006	259-263	CK2 PHOSPHO_SITE	POC000006
PS00006	286-290	CK2 PHOSPHO_SITE	POC000006
PS00006	295-299	CK2 PHOSPHO_SITE	POC000006
PS00006	302-303	CK2 PHOSPHO_SITE	POC000006
PS00006	317-321	CK2 PHOSPHO_SITE	POC000006
PS00006	336-340	CK2 PHOSPHO_SITE	POC000006
PS00006	342-343	CK2 PHOSPHO_SITE	POC000006
PS00006	372-376	CK2 PHOSPHO_SITE	POC000006
PS00006	427-431	CK2 PHOSPHO_SITE	POC000006
PS00006	447-451	CK2 PHOSPHO_SITE	POC000006
PS00006	462-463	CK2 PHOSPHO_SITE	POC000006
PS00006	522-526	CK2 PHOSPHO_SITE	POC000006
PS00006	597-601	CK2 PHOSPHO_SITE	POC000006
PS00007	19-191	MYRISTYL	POC000007
PS00008	207-213	MYRISTYL	POC000008

```

HMM_NAME      TNFR/NGFR cysteine-rich region

HMM            *CpeGtYtDWNHvpqClpCtrCsPEMGQYMvqPCTwTQNTVC*
               C+ ++ + N+ ++ + ++ + ++ ++ ++ VC
Query          274 CESNQPYG-NT-RITFEEDQVDS--TLIDSSSHDEWLDAVC    310

```

DNF2phtes3.7j3

group: cell cycle

DNF2phtes3.7j3.2 encodes a novel 628 amino acid putative protein kinase, which is related to the C-TAK1/Cdc25C associated protein kinase.

Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to serine/threonine-specific protein kinases

complete cDNA, complete cds, potential start at bp 128, few EST hits

Sequenced by BHF2

Locus: unknown

Insert length: 3443 bp
Poly A stretch at pos. 3399, polyadenylation signal at pos. 3376

```
1  GTGCTTACT  CGCGGCTCTG  CTACTGCTGT  GGCCTCCGGT  CCGTGCGGG
51  GACCTTGCCG  CGCGGCTTCA  GCCTCCCGCG  CACAGCTTAC  TGATTCCTCT
101  CGCGGCTTGG  CTCACCTCCT  GCTCGCATG  GAGTGCCTGG  TTTTCGGCGG
151  GCCTCCCGCG  CCACTCCTCT  CGCGGCGAGA  GCTAGCCCGG  CCGCTGGCGG
201  AAGGCTGAT  CAATGCGGCG  AGGCTCTTAT  TGAGAGGCA  GGGGTGAG
251  CGGACACACC  ACAGGCACAA  CCGCGGCGAC  CGCTACGAGT  TCTTGAGAC
301  CCGGCGCAAA  GGCACCTACG  GGAAGGTGAA  GAAAGCGCGG  GAGAGCTCGG
351  GCGGCTGCTG  GGCCTACGAG  TCAATCCGGA  AGGACAAAT  CAAGATGAG
401  CAAGATCTGA  TGCACATACG  GAGGGAGATT  GAGATCATGT  CATCACTCAA
451  CCACGCTCAC  ATCAATGCGA  TCCATGAAGT  GTTTGAGAAC  AGCAGCAAGA
501  TCTGTATCTG  CATGAGATAT  CCGAGCCGGG  GCGACCTTTA  TCACTACATC
551  AGCGAGCGCG  ACCACTCTAG  TGAGCGCGAA  GTTAGGGATT  TCTTCGGCA
601  GATCGTCTCT  GCGCTGCACT  ATTGCAATCA  GAACAGAGTT  GTCCACCGAG
651  ATCTCAAGCT  GGAGAACATC  CTCCTTGATG  CCAATGGGAA  TATCAGAGAT
701  GCTGACTTGG  GCGCTTCCAA  CTTCTACAT  CAAGGCAAGT  TCTTGAGAC
751  ATTCTGTGGG  AGCCGCTCT  ATGCTCGCG  AGAGATTGTC  AATGGGAAGC
801  CCTACACAGG  CCGAGAGGTG  GACAGCTGGT  CCTGGGTTGT  TCTCCTTAC
851  ATCTGTGTCG  ATGACACCAT  GCCTTTGAT  GGCATGAC  ATAGATCTCT
901  ACTGAAACAG  ATCAGCAAGG  GGGCTACGG  GGAGCCACT  AAACCTCTCT
951  ATGCTGTGGG  CCGTATCGGG  TGCTTTTGA  TGGTGAACCC  CACCGCGCGG
1001  GCGACGCTGG  AGGATGTGGG  CAGTCACTGG  TGGTCAACT  GGGGCTACGC
1051  CACCGCACTG  GAGAGCGAGG  AGCTTCCGA  TGAGGTTGG  CAGCTTGGCA
1101  GTGACTCTGG  CGCGGCTTCC  ATGGCTGACT  GCGTCCGGCG  TTCCTCCCGG
1151  CCGCTCTGGG  AGAATGGGCG  CAGGTTGTC  AGCTTCTTCA  AGCAGCATGC
1201  ACTGTGTGGG  GGAAGACCA  CCGCTGCTT  GAGCGCGAG  CATCTCTCA
1251  AGAATCTCCG  CAAGGAGAA  GACATGGCC  AGTCTCTCCA  CAGTGACAGC
1301  GCTGATGACA  CTGCGCATGG  CCGTGGCAAG  AGCAACTCCA  AGCTGCCAAA
1351  GCGGATCTCT  AGAGAGAGG  TGTGAGCTG  TCGAGAGGG  GTACAGAGG
1401  ACCCTCCGGA  GCTCAGCCCA  ATCCCTGCG  GCGCAGGCA  GGTGCGCCCG
1451  CTGCTCCCGA  AGAAGGCAAT  TCTCAGAG  CCGCGACAG  CGAGTCTGG
1501  CTACTACTCT  TCTCCGAGC  CAGTGAATG  TGGGAGCTC  TTGGAGCGAG
1551  GCGACTGTGT  TGTGATGGG  GATCCGAGG  AGCAGAGCC  TCGCGAGCT
1601  TCAGGGCTGG  TCTCTCATG  CAAAGCATC  CTCAACTCA  ATGCGAAGTT
1651  CTCCGAGACA  GCTTGGAGG  TCGCGGCCCG  CACCACTTC  GCTCCTCTGG
1701  ATCAACTGCG  CCGACTGCG  CCGTGGCGG  GCGCACAGC  ACCTCTAGGG
1751  GCTGTGAGCG  AGGACAGAT  CTTCTCTCT  GAGTCTTTTG  ACCAGCTGGA
1801  CTGCGCTGAA  CGGCTCCAG  AGCCGCACT  GCGGGGCTGT  GTGTCTGTGG
1851  ACAGCTTAC  GCGGCTTGA  GAGCCGCTT  CAGAGGGCG  TGAAGCTGC
1901  CTGAGGCGCT  GCGCGCAGG  TCTTTGGGG  GACAGCTGCT  TTTCCCTGAC
1951  AGACTGCCAG  GAGGTGACG  CGACTACCG  ACAGGACTG  AGGCTCTGCT
2001  CAAAGCTCAC  CTGAGTGAG  TAGGCAATG  CCGAGCCCG  TCAGGCTCTC
2051  AGATGAGCT  GCTTGCGCC  CGAGGAGGA  TCTCTCTTC  CCGACTCTCC
2101  AGACCTGCA  TCCAGCTCA  GAAGCTGAG  AGGTTTGA  CTGGAGCCCT
2151  GAGCAGGCT  GATATGGGA  AGTAGGCAA  TGAATGCGC  CAGGGTTTCA
2201  GTTCTGTCT  TCAGCTCTG  TGAGCAAGA  GATATCAA  GAGAGGGAA
2251  CGGGAATGCC  CGCGACAGG  TCCACATTC  CTGTTTCTG  TGTACATGGG
2301  GGGGCGACAG  AGACCTGAA  AGAGACTCT  CCGAGGCGC  ATCTCTGCA
2351  TCCATGAA  ACTCTGACA  CATGTGCTT  TCTAGACA  GCTCTTCC
2401  TACTCATTC  CTCCCAAGT  GGGGCGAGC  CTCCTTAC  ACACATTC
2451  GTTCTTACCA  ACCACGAA  CTGATGTT  GACCCCTAA  TGTGATGAG
2501  GCATCTCTGG  AATGCTCTG  AGTAACGCT  CTTATTTT  ATTTTATTT
```


WO 01/12659

PCT/IB00/01496

```
2551 TTATTTATTT ATTATTTTT TTGAGACGGA GTTTCGGCTT TGGTGCCGAG
2601 GCTAGAGTGC AATGGCGCGA TCTCAGTCA CCTCAACCTC GGCCTCCGGG
2651 GTTCAGCGCA TTCTCTGCCC TCAGCCTCCC TAGTAGCTGG GATTACAGGC
2701 GCGCGCCACC ATGCGCGGCT AATTTTGTAT TTTTAGTAGA GACAGGOTTT
2751 CTCAGTGTTC GTCAGGCTGG TCTCAAACTC CGACCTTCAG GTGATCAGC
2801 CACCTCGGCC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCGCGCCC
2851 CACCTAACCC TTCTTATTT AGCCTAGGAG TAGAGAGACA CAATCTCTGT
2901 TCTTCAATG GTTCTCTTCC CTTTTCATC CTCGAAACCT GCGCTGAGCC
2951 TCTGAAGTT GCTGCTGTGA ATCTGAAGA CTGAAAGC CTCGCGCTGC
3001 TGTGTGACT TCATCTCAAG GCGGCCAGCC TCTCTGGAC TCCACTTGG
3051 ACCTCAGTCA CTCAGAACTT CTGCTCTTAA GCTGCTTAA ACTCCAGCT
3101 ATGATCTCT TCTCTAGGCC TTAGGACTC TAGAATGCC ATATTTATTT
3151 TTATGTTCTT GCGTTTGTCT TTAGGAAGA GTGAATCTG CTGTTTCAA
3201 TAATGTGAAT GCTATGTTCT GGGAAATCCC ACTATGACAT CTAGTTTTC
3251 TGTAGAGAGA GATATTTTTC CAATATTCT CACTCTCTCC CAGAACGCC
3301 CACACTCCAC TCCACTCTT TGAATCTCT TACCTAATGG TCTCTACCTA
3351 ATGGACCTCC GTGGCAAAA AGTACCATTA AAACCGAAA GGTGATTGGA
3401 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

98202387:
C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and
promotes 14-3-3
protein binding.

Peptide information for frame 2

ORF from 128 bp to 2011 bp; peptide length: 628
Category: strong similarity to known protein

```
1 MESLVFARRS GPTPSAAELA RPLAEGLIKS KPPLMKQAV KRHHKHNLK
51 HRYEFLFETLG KGTGKVRKA RESSGRLVAI KSIRKDKIK DDLAMHRRK
101 IELHSLAHNP HIALHEVIE KSKIVIVNE YASRGDTOT ISEFQLSER
151 EAKHFFROIY SAVHYCHQNR VVRKDLKLEH ILDANGNIK IADFLSLMLY
201 HOGKFLQTFK GSPLVASPEI VNGKPTTQPE VDSMLGLVLL YILVHGTNPF
251 GGDHRIKLVK QISNGAYNEP RFFSDAGCLI RMLLAWVPTT RATLEDVASH
301 WVMVGVYATR VGEQAPHEG GHPGSDSARA SKADWLRRSS RPLENGAKV
351 CSFFKHAPG GGSFTPLGLR QHSLKSKKE NMAQSLHSD TADDTAHRPG
401 KSKLRLPKGI LKRVVASAE CYGDEPFELS PIPASPGDAA PLLPKGILK
451 RPNQRESQIV SSEPFESGE LLDAGDVFS GDFKQKPPQ ASGLLLHRKG
501 ILKNGKFSQ TALELAAPT FGLDELAPF RFLARSRPS GAVSEDSILS
551 SESFDQLDLP ERLPEPLAG CYSVDLITGL EEPSPGPGS CLRWRRQDFL
601 GDSFSLTDC QGVATVRA LRVCSKLT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for ORFsphtes3_7j3, frame 2

No Alert BLASTP hits found

Peptide information for ORFsphtes3_7j3, frame 2

Report for ORFsphtes3_7j3.2

(LENGTH) 628
(MW) 69612.39
(pI) 9.01
(RCOML) TREMBL:AB011109.1 gene: "KIAA0537"; product: "KIAA0537 protein"; Homo sapiens
NR04 for KIAA0537 protein, complete cds, 1e-152
(FUNCAT) 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
5e-66
(FUNCAT) 11.01 stress response [S. cerevisiae, YDR477w] 5e-66

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 5e-66
[FUNCAT] 90 classification not yet clear-cut [S. cerevisiae, YLR056w] 6e-54
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YLR056w] 6e-54
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 9e-51
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGL101w] 9e-51
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-45
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.22.01 cell cycle checkpoint proteins [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YNR001c] 2e-42
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-34
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL139w CTK1 - carboxy-terminal domain] 2e-28
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-28
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 4e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 03.13 meiosis [S. cerevisiae, YDR523c] 8e-22
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 6e-21
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, fatty acylation and processing) [S. cerevisiae, YPL031c] 6e-21
[FUNCAT] 10.05.09 regulation of p-protein activity [S. cerevisiae, YBL016w] 7e-19
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YDL159w] 3e-18
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-17
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 4e-16
[FUNCAT] 04.03.99 other trans-transcription activities [S. cerevisiae, YOR061w] 1e-15
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w] 5e-15
[FUNCAT] c energy conversion [M. genitalium, MG169] 3e-12
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.07 organization of endoplasmic reticulum [S. cerevisiae, YHR079c] 8e-05
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 8e-05
[BLOCKS] BL00479C Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS] BL00219B Receptor tyrosine kinase class II proteins
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[SCOP] dfgk_ 5.1.1.1.1.9 MAP kinase Erk2 (rat Rattus norvegicus) 1e-77
[SCOP] dlwfc_ 5.1.1.1.1.8 MAP kinase p38 (human (Homo sapiens)) 4e-68
[SCOP] dlksa_ 5.1.1.1.1.7 (1-350) Twitchin, kinase domain (Caenorhabditis) 2e-85
[SCOP] dlkba_ 5.1.1.1.1.6 Twitchin, kinase domain (california sea har) 1e-80
[SCOP] dlphk_ 5.1.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 2e-76
[SCOP] dlirk_ 5.1.1.1.2.4 insulin receptor (human (Homo sapiens)) 1e-69
[SCOP] dlqpm_ 5.1.1.1.1.4 cAMP-dependent PK, catalytic subunit (mouse (Mus)) 1e-84
[SCOP] dlfgka_ 5.1.1.1.2.3 Fibroblast growth factor receptor 1 (human (Homo sapiens)) 1e-68
[SCOP] dlydra_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit (bovine (Bos)) 9e-85
[SCOP] dlfaa_ 5.1.1.1.2.2 (168-437) c-src tyrosine kinase (human (Homo sapiens)) 1e-69
[SCOP] dlcdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit (pig (Sus)) 1e-85
[SCOP] dhcka3_ 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck (human (Homo sapiens)) 5e-66
[SCOP] dlcan_ 5.1.1.1.1.1 Casein kinase-1, CK1 (Schizosaccharomyces pombe) 9e-47
[SCOP] dljaue_ 5.1.1.1.1.1 Cyclin-dependent PK (human (Homo sapiens)) 1e-75
[SCOP] dlckja_ 5.1.1.1.1.10 Casein kinase-1, CK1 (rat (Rattus norvegicus)) 5e-54
[EC] 2.7.1.38 Phosphorylase kinase 1e-36
[EC] 2.7.1.123 G2/caldesmon-dependent protein kinase 4e-40

[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 1e-61
 [EC] 2.7.1.117 Myosin-light-chain kinase 2e-40
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 1e-61
 [EC] 2.7.1.37 Protein kinase 7e-42
 [PIRKE] phosphotransferase 6e-66
 [PIRKE] nucleus 1e-64
 [PIRKE] calcium 7e-35
 [PIRKE] duplication 1e-38
 [PIRKE] tandem repeat 4e-39
 [PIRKE] phorbol ester binding 1e-38
 [PIRKE] zinc 1e-38
 [PIRKE] cell cycle control 1e-42
 [PIRKE] serine/threonine-specific protein kinase 8e-68
 [PIRKE] oncogene 1e-40
 [PIRKE] phospholipid binding 1e-38
 [PIRKE] autophosphorylation 1e-64
 [PIRKE] brain 1e-40
 [PIRKE] heterotrimer 2e-36
 [PIRKE] mitosis 7e-42
 [PIRKE] polymer 1e-35
 [PIRKE] magnesium 6e-66
 [PIRKE] ATP 8e-68
 [PIRKE] polyprotein 1e-40
 [PIRKE] phosphoprotein 1e-64
 [PIRKE] apoptosis 4e-39
 [PIRKE] glycoprotein 7e-42
 [PIRKE] leucine zipper 3e-35
 [PIRKE] skeletal muscle 7e-35
 [PIRKE] protein kinase 5e-41
 [PIRKE] cAMP binding 3e-38
 [PIRKE] testis 9e-36
 [PIRKE] purine nucleotide binding 2e-49
 [PIRKE] calcium binding 8e-39
 [PIRKE] alternative splicing 3e-37
 [PIRKE] P-loop 2e-49
 [PIRKE] lipoprotein 2e-33
 [PIRKE] segmentation 1e-33
 [PIRKE] core protein 1e-40
 [PIRKE] muscle 7e-35
 [PIRKE] myristylation 2e-33
 [PIRKE] EF hand 8e-39
 [PIRKE] cell division 2e-40
 [PIRKE] calmodulin binding 4e-40
 [SUPFAM] ribosomal protein S6 kinase II 5e-36
 [SUPFAM] fibronectin type III repeat homology 3e-33
 [SUPFAM] immunoglobulin homology 3e-33
 [SUPFAM] calcium-dependent protein kinase 8e-39
 [SUPFAM] AMP-activated protein kinase 6e-66
 [SUPFAM] protein kinase akt 3e-42
 [SUPFAM] protein kinase SRK1 1e-42
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-37
 [SUPFAM] calmodulin repeat homology 8e-39
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 6e-33
 [SUPFAM] protein kinase C zeta 1e-36
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-34
 [SUPFAM] death-associated protein kinase 4e-39
 [SUPFAM] pleckstrin repeat homology 3e-42
 [SUPFAM] ankyrin repeat homology 4e-39
 [SUPFAM] protein kinase homology 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-41
 [SUPFAM] protein kinase C zeta-binding repeat homology 1e-38
 [SUPFAM] twitchin 3e-33
 [SUPFAM] protein kinase C delta 1e-38
 [SUPFAM] cAMP-dependent protein kinase 6e-33
 [SUPFAM] protein kinase cdrl 7e-42
 [SUPFAM] protein kinase C C2 region homology 3e-37
 [SUPFAM] protein kinase C alpha 3e-37
 [SUPFAM] yeast protein kinase C 5e-36
 [SUPFAM] kinase-related transforming protein 1e-41
 [SUPFAM] kinase interaction domain homology 1e-42
 [SUPFAM] Gsp-akt polyprotein 1e-40
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 4e-40
 [SUPFAM] protein kinase C mu 4e-33
 [PROSITE] PROTEIN_KINASE_ATP 2
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 4
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 12

```
(PROSITE)    ASN_GLYCOSYLATION      2
(PROSITE)    PROTEIN_KINASE_ST      1
(PFAM)       Eukaryotic protein kinase domain
(KW)         All_Alpha
(KW)         30
(KW)         LOW_COMPLEXITY    10.51 %

SEQ    MESLVFARRSGPTPSAAELARPLAEGLIKSPFLMKKQAVKHHKHNLRHYEFLETIG
SEG    .....XXXXXXXXXXXX
lc1pE  .....HKKHHHHHHHHHHHCCCCCCCC--GGGEEEEEEE

SEQ    KGYGKVKKARESSGRLVAKSIKDKIKDEQDLMIHREIETMSSLNHPHIIAIEVFE
SEG    .....
lc1pE  .....CTTTEEEEEETTTTKEEEELKHHHHHHHHHHHHHHHHHHHCCCTTTTCCCEEEEE

SEQ    NSSKIVIVNEYASRGDLVDYISENQQLSEREARHFFQIVSAVHYCHQNRVVRDLALEH
SEG    .....
lc1pE  .....ETTEEEEEECTTTTBNHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCECCCCGGG

SEQ    ILLDANGHRIADPGLSLHYDQFLQTFGSGPLVYASFEIYNGKPYTGPVEDSWSLGVL
SEG    .....
lc1pE  .....EEETTTTCEEECTTTTET-TT-BCCCCCGGGCHHHHHCCCCB-HHHHHHHHHHH

SEQ    YILVHGTNFPDGHDKILVKQISNGAYREPPPSDACGLRWLMVNPTRATLEDVASH
SEG    .....
lc1pE  .....HHHHHCCCTTTTTHHHHHHHHHHCCCCCTTCHHHHHHHHHHTTTTGGGTTTHHHHHC

SEQ    WVMNGYATRVGEQAPREGGHPGSDSARASADNLARSSRPLENGAKVCSFFQKAPG
SEG    .....
lc1pE  .....GG

SEQ    GGSITPGLERQHSLLKSRKENDMAQSLHSOTADTAHRPGKSHLKLPGILKKKVSASAE
SEG    .....
lc1pE  .....

SEQ    GVQEDPFELSPFASPGQAAPLLPKYGLIETPQRESGVYSSPEPSGELLADGVYS
SEG    .....XXXXXXXXXXXX
lc1pE  .....

SEQ    GDFKEQKPPQASGLLLKRGILKMGKFSQALELAAPTTFGLDELAPPPLARASRPS
SEG    .....XXXXXXXXXXXX
lc1pE  .....

SEQ    GAVSEDSILSSESFDQLDLPALPEPPLRGCVSDVNLGLEEPPSEGPGLARWRQDPL
SEG    .....XXXXXXXXXXXX
lc1pE  .....

SEQ    GDSCFSLTDCQEVATATYQALRVCSKLT
SEG    .....
lc1pE  .....
```

Prosite for DKFiphtes3_tj3.2

PS00001	121->125	ASN_GLYCOSYLATION	PD0C00001
PS00001	576->580	ASN_GLYCOSYLATION	PD0C00001
PS00004	290->294	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	337->341	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	413->417	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	30->33	PKC_PHOSPHO_SITE	PD0C00005
PS00005	74->77	PKC_PHOSPHO_SITE	PD0C00005
PS00005	82->85	PKC_PHOSPHO_SITE	PD0C00005
PS00005	122->125	PKC_PHOSPHO_SITE	PD0C00005
PS00005	142->145	PKC_PHOSPHO_SITE	PD0C00005
PS00005	148->151	PKC_PHOSPHO_SITE	PD0C00005
PS00005	289->292	PKC_PHOSPHO_SITE	PD0C00005
PS00005	327->330	PKC_PHOSPHO_SITE	PD0C00005
PS00005	339->342	PKC_PHOSPHO_SITE	PD0C00005
PS00005	373->376	PKC_PHOSPHO_SITE	PD0C00005
PS00005	377->380	PKC_PHOSPHO_SITE	PD0C00005
PS00005	616->619	PKC_PHOSPHO_SITE	PD0C00005
PS00006	13->19	CK2_PHOSPHO_SITE	PD0C00006
PS00006	133->137	CK2_PHOSPHO_SITE	PD0C00006
PS00006	148->152	CK2_PHOSPHO_SITE	PD0C00006
PS00006	227->231	CK2_PHOSPHO_SITE	PD0C00006
PS00006	293->297	CK2_PHOSPHO_SITE	PD0C00006
PS00006	331->335	CK2_PHOSPHO_SITE	PD0C00006
PS00006	377->381	CK2_PHOSPHO_SITE	PD0C00006
PS00006	391->395	CK2_PHOSPHO_SITE	PD0C00006

PS00006	461->465	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->582	CK2_PHOSPHO_SITE	PDOC00006
PS00006	606->610	CK2_PHOSPHO_SITE	PDOC00006
PS00007	453->460	TYR_PHOSPHO_SITE	PDOC00007
PS00007	453->461	TYR_PHOSPHO_SITE	PDOC00007
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	360->366	MYRISTYL	PDOC00008
PS00016	134->137	RGD	PDOC00016
PS00107	59->82	PROTEIN_KINASE_ATP	PDOC00100
PS00107	59->86	PROTEIN_KINASE_ATP	PDOC00100
PS00108	171->184	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZp468j3.2

HMM_NAME	Eukaryotic protein kinase domain
HMM	YelgRlIcGcFGLVYVCLMGTGcVIAIKIikhsms.....FIREI YE+***G+G+G+V+K+**+G+VAIK I+K+***+**REI
Query	53 YEFLETLGKGTGKVKKARSSGRLVAIKSIRKDKIKDEQDLHIRREI 101
HMM	qINR+LNNP+II+FE+**I+**NEY+GDL+DYI+***SE+ +IM+LNNP+II+FE+**I+**NEY+GDL+DYI+***SE+
Query	102 EIMSLNHPHIIAINEVFE-NSSKIVIVNEYASGDLVDYISERQQLSER 150
HMM	eTcfINyQILrGMeYLISNqIIRDLKPEMILIdNgqTKicDFGLAqM E+R+***QI+***Y+R+***HRDLK ENIL+D NG+IKI+DFGL+**
Query	151 EARHFFRQIVSAVYCHQNRVVRDLKLEMLLDANGHRIADFGLSHLY 200
HMM	nnYcMtlfcGTpWYNAFEVIlmg.nyYtlLVDMMSFGILMENDNTGep + +**TFCG+P Y +PE+ +G +Y +**VD MS+G+L+***G+
Query	201 HQGFLQTTCGSPLYA-SPEI-VNCRPTGPEVDMSLGVLLTILVHCTH 248
HMM	PfYddnMemImriIq:frpDpncSeEIyDfNwCWnyDPeLRPTf+QI PF+**+ I +***+P S+ +**RM+**P+R T +**
Query	249 PFDGNDKILVKQISNGATREPPKPSD-ACGLIRLLMNPTRATLIDV 297
HMM	LNNP+ H M+
Query	298 ASHMMV 303

WO 01/12659

PCT/IB00/01496

DKFSphtes3_7j8

group: testes derived

DKFSphtes3_7j8 encodes a novel 410 amino acid protein nearly identical to human WUGSC:H_DJ1159004.1.

The novel protein contains an additional C-terminal domain, which is not present in WUGSC:H_DJ1159004.1. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_DJ1159004.1 similarity to YBL104p

verifies and extends the genmodel WUGSC:H_DJ1159004.1 similarity to S.cerevisiae YBL104p

Sequenced by BMFZ

Locus: /map="p21-p22"

Insert length: 1353 bp
Poly A stretch at pos. 3231, no polyadenylation signal found

```

1 GCAGAAATATG TTGTATTGT GGCATAGTC ATATTACAC TATCATAAAA
51 TTATGGCCGA GAAGTAAAT ATTCTAAATG TGTCACATA GTTCTCTGTA
101 AAATGACCTT ATTTCCAAA TATATTTTGA AATAAACAA TATAAAAAATG
151 TTTCCTGTTT TTAGGATGG TGGAAAGCG CAGACATAT TGGAGTGGGT
201 TGGATAGGCA AAGTGATATT CAAATTTTAA ATGAAGAGAG AATCTTAGCT
251 TTACAGCTTT GTGGTGGAT AAAGAAGGA ACGGATGTAG ACOTGGGGCC
301 ATTTTGAAC TCCTCTGTAC AAGAAGGGA ATGGCAAGA GCTGCTGCTG
351 TGGCATTTCT CACTTGGAT ATTGGCGAG CAATCAAT CTGAATGAA
401 GGGGCATCTT CTGAAAAGG AGATCTGAAT CTCAATGGG TAGCAATGGC
451 TTTATCGGGT TATACGGATG AGAAGAAGCT CTTTGGAGA GAAATGTGTA
501 GCACACTGCG ATTACAGCTA AATACCGCT ATTGTGTGT CATTTTGGCA
551 TTCTGCACAA GTGAACAGG ATCTTAGAT GGAGTTTGT ATGAACAA
601 AGTTGCAGTA CGTCACAGAG TGGCATTTGC TTGTAAATTC CTTAGTGATA
651 CTCAGTTAAH TACATACAT CAAAGTTTGA CCAATGAAT GAAGAGGCT
701 GCAAAATTTG AGGAATTTT GCTTACAGGC CTACTAAG ATGAGTGG
751 CTTAATGGAG AGTTATGTT ATAGAAGTGG AGATOTCAA ACAGCAAGTT
801 ACTGTATGTT ACAGGTTTCA CTTTATAGT TCTTAAAGA TGAAGGGTT
851 CAGTACTGGA TTGCAATTA TAGAAATTA TTAGATGCT GGAGCTTTTG
901 GCATAAAGCA GCTGAATTTG ATATTCAGG GACTAAGTG GATCCAGTT
951 CCAAGCCTTT AGCACAAGTT TTGTGAGTT GCAATTTCTG TGCCAAGTCA
1001 ATCTCTACA GCTCTTACG TTGCTCTAT CAGGCGAGG GTTTATGCA
1051 GTATGGTGTG AGTGGCTCAC CAACGAATC TAAAGTCACA AGTTCTCTG
1101 GCTGTGAAA ACCACTTCTT CGATGTGGG TTGTCTCAT TAATATGGGA
1151 ACACAGCTT CTAGCTCTCC TGGAGAAC CCATCAGAT AAAAGTGA
1201 CTTGAGCAAG GACAAAAAT TAGCCCAAT TAACAAGTG TTACATGGT
1251 GTCATAATG GAGGACGGT GGACATCTG GACATATGCT TACTTGTTTC
1301 AGGACACATG CAGAGTGGCC TTGTCTGCA TGCAGTGTA AATATATGA
1351 CTTGATACA ACGGGAATC TGTACTCTG AGAAGCTCT CAGCATAAA
1401 ATCTTACCAC CTTAAGAGAA CCTTCAAGT GTGGAGCTT CTAGTAGGTG
1451 TCTTTCATAG CTCAGAACCA TACTTCAGAA CAGGCTATC ATGACTTACC
1501 TGTATGGGA AAATTAATCA TCTATCAGA TCAGCAGTT TGATTTTGA
1551 GGTATTTTGA TATGCTTAC AGAGACAAAT GTGCCAATA TAAACATGGA
1601 AGTATAGACA TGAGTTCTGT TCAGCAGTT GAAAAGTCTG ATTTAGAAA
1651 ACTTCTAAG TTTTGGTGA AATATGAC ACTCTAGAG CAGATTTCT
1701 GGAGAGCCA AGAACAGCT TTGAGCTAT ATCTTCAAG CTGAAGCTGG
1751 ATATCTTTCA ATAAATATG TGACTTTTA AATAAAATG ACTAATCTG
1801 TGATTACAGC AATAGTTTTA AGTTCAGCT TGCTTAAAT TCTTACAGT
1851 TAATTAAAG TTATGATTT TACTTTTAG AATTGCAGG CCGTATGCC
1901 ACACGGAGA ATATTTTFA TACTGTCTG TTATATATG GTCTATGTGT
1951 GTGTGTATAT TTATGTGTG ATGTATAAT ATGTACTTT TAAAGAGGC
2001 TTTTGGCTC TTGATTTTA AGTAAGCAA TCTTTGCA TAACTATATC
2051 GTCTTCTAG AAAAGCCAAG ATGAAGATC TATCTTACAA CTTTCTCT
2101 TCAATAGAGA AAAACATGTA CCAITTCAGG TGAACATACA AAATTTTCAC
2151 TTCTTACTT TGGCTTCCA ATGCTGAT TTGCTTCAA AGGTTTCT
2201 CCAATTAAT TTGTCACTT ATCTTCATCA CTTGAGACA TTTTACTGCA
2251 TACAAGTCT ATGCAAGATT ATATTAAT AGCCATTAG TATAATCTAT
2301 GTCACTGTT CTGTCTGTC AATCTGCTC CTGATTTGA ATACCAATCC
2351 TTCTTCTTC CAGGTAGAC TGGACACTT TGGCAATA GGCTCACTTC
2401 AGAGACCAT TTACATGTAA GTTTTAAAT TGAAGTGTTA CTGGGGCTAA
2451 GTCAGGCAT TTATTTAAAA CATTTTTCT TCTCTATTG ATAGCTAGT
2501 AGTTGTAAAG GAAATACAAA GAATTTACAA GATGCTTCTG TGTCTCTG

```

```
2551 CGTATGCAGA GGGACTGAC TAGGAATTTT GTAGTTGAAG CTGTGTCAT
2601 AAGAGCTTAA TCTTATTTA TAGATTTGG AGAATAAATA CAGAAATTTT
2651 AAGAGCTTTC GTATTAGCAG TTTTGCCCTA TAAAAACTAA GATTGTGAG
2701 ATTACTTTGA GGTGTAACCT AAKATATATA ACTGATTTAA ATTATTTTTT
2751 TACCTTGAGT GTCTGATACA TAAACCCCTT TCTAGGAAA ACATTGGAG
2801 TAGTACATAT TTACTCTAAA TGCTCACCT GCATGACAGT CTTTCAAAAT
2851 GAAGACATAT GTAAATGAAA TTTTCTTTTA AGATTTGCTA TTAAGGGTAC
2901 TTTTCCAGC CTTCATTGA GTAAATCTTA ATTGATTICA TTTTATTAAC
2951 ATATACCCCT TACCTTTAT ATTTCATTG AGGTGTTCTT TCAAACTTA
3001 CTGCTTTAAA TATGAAGTC AGCTTTAAT AATGTCAGAC TCATATGAT
3051 TTCAATCTC ATTACTTAA GTAAATGTA AATTATCTC AATAGTTAC
3101 AAGTTTGGG AATACAGAT AAAACATGA TGTAAAGTCT ATTATGTAT
3151 ATGCTTATTT GTAACTTAA TATATGAGG TACATTTT AGATTGTAT
3201 GTATGTGTA ACCTTTTAA TGTTTCTGT GAAAAAATA AAAAAAATA
3251 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
3301 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
3351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 167 bp to 1396 bp; peptide length: 410
Category: known protein
Classification: unclassified

```
1 MVESSRNHWS GLDKQSDION LNEERILALQ LCGMIRKQTO VDVPFPLNSL
51 VQEGEMERAA AVALFNLDIR RAIQILNEGA SSEKGLNLN VVAHALSGYT
101 DEKNSLWREH CSTLRQLQWN PYLCVNFAPL TSETGSDGV LIENKVAVRD
151 RVAFACKTLE DYOLHWYIEK LTHWKEACH LQILLTGLT KQGVLMSEY
201 VORTQDVQTA SYCHLQSPFL DVLKDERVOY WIENYRNLLD AWRFMKRAE
251 FDIRSKLQSP SSKPLAQIVY SCHFCGESIS VCSAVPHQG RGFSDYGVSG
301 SPYKRVKTSK PCKRFLPKK ACCLINWQTP VESCPGQTKS DEVLVLSKQ
351 ELADFNWFTT MCHCRHUGH AGHKLWFRD HAZCPVSACT CRCHQLDTTG
401 NLVPAETVOP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3.7j8, frame 2

PIR:S45391 probable membrane protein YBL104c - yeast (Saccharomyces cerevisiae), N = 2, Score = 416, P = 4.5e-47

TREMBL:AC004982.1 gene: "MUGSC:H_DJ1159004.1"; Homo sapiens PAC clone DJ1159004 from 7p21-p22, complete sequence., N = 1, Score = 2038, P = 7.6e-211

>TREMBL:AC004982.1 gene: "MUGSC:H_DJ1159004.1"; Homo sapiens PAC clone DJ1159004 from 7p21-p22, complete sequence.
Length = 379

HSPs:

Score = 2038 (305.8 bits), Expect = 7.6e-211, P = 7.6e-211
Identities = 379/379 (100%), Positives = 379/379 (100%)

```
Query: 1 MVESSRNHWSGLDKQSDIONLNEERILALQQLCGMIRKQTOVDVPFPLNSLVQEGEMERAA 60
      1 MVESSRNHWSGLDKQSDIONLNEERILALQQLCGMIRKQTOVDVPFPLNSLVQEGEMERAA
Sbjct: 1 MVESSRNHWSGLDKQSDIONLNEERILALQQLCGMIRKQTOVDVPFPLNSLVQEGEMERAA 60

Query: 61 AVALFNLDIRRAIQILNEGASSEKGLNLNVVAHALSGYTDKNSLWREKSTLRQLQWN 120
      61 AVALFNLDIRRAIQILNEGASSEKGLNLNVVAHALSGYTDKNSLWREKSTLRQLQWN
Sbjct: 61 AVALFNLDIRRAIQILNEGASSEKGLNLNVVAHALSGYTDKNSLWREKSTLRQLQWN 120
```

Report for DKFZphtes3 718.2

[illegible]

(No Pfam data available for DKFZphes3 718.2)

WO 01/12659

PCT/IB00/01496

DKFphtes3_7p10

group: Cell Cycle

DKFphtes3_7p10.1 encodes a novel 422 amino acid putative protein, which is closely related to the *Xenopus laevis* XPMC2 protein.

In fission yeast the kinases Wee1 and Mik1 control that initiation of mitosis starts after completion of DNA synthesis. Yeast in which both Wee1 and Mik1 kinases are defective exhibit a mitotic catastrophe phenotype. XPMC2 of *Xenopus* rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function. The XPMC2 protein is localised in the nucleus in *Xenopus* oocytes. The new protein is the human orthologue of this gene.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to XPMC2 protein

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="9q34"

Insert length: 2380 bp
Poly A stretch at pos. 2341, polyadenylation signal at pos. 2318

```

1  AGCCTGCGTG  CTGAGGTATG  CGCAACGCGT  GCGGGGTCTC  TTCCGAGTC
51  TTTTCTCGGA  CGGATCTCTT  GCGGTGGGTT  TTTTTCGGGC  TGGCTTGGGC
101  AGCGCGCTTT  GCTGCCAGGG  CGCGGGGCCG  GGGAGGCGCG  GGGTCTCGGG
151  TGGCGCGCGG  CCGAGCGGCT  GGACGGCAGC  AGGATGGGGA  AGGCCAAGGT
201  CCCCCTCTCC  AAGCGCGGCC  CGAGCAGGCC  CTTGGCTTAG  CCGGTCTCTG
251  TCAGAGCGCT  CACTGGGAG  AAAAACAAGA  AGAAAAAAG  GTTTTGGAAA
301  AGCAAGCGCG  GGAAGTAG  CAGAAGCCA  GCAAGCGGCC  CCGGTCTCTG
351  GGTGCGACCT  CCAAGCGCAG  CAGAAGCTT  TTCTCAAAK  TCGAGGGCG
401  TCGAAGCTG  CTTCTGA  CAATAATCT  AGGCTCCAGA  AANSGCTCTT
451  GTATCTCTC  AGATGGGTT  CAATAAGAG  CCAAAATTA  TCCACAAAA
501  CAAAAAAGAG  ACCTCGCTC  AAGTGAAGG  AGAGAGATG  CCGGACAGAA
551  AAGACCGGA  GCGCGCGG  GCTCTCTTC  CTTAGGTT  CAGATGGAC
601  AGAGGCGCG  CACTACCTC  CACCAAGCG  ACTGGAACG  AGCACAATA
701  ACATCGAGA  TACAGAGG  AAGCTAAG  AGCGAGCG  AGCCGACCC
751  ACCGAGGA  ACATCTGTT  TGACGAGCT  GACCCAGCG  ATATCGAAG
801  TGCCATAGT  CCAGAGCGG  CCAAGTAG  GAGGAACAG  TTGGTTCAG
851  GCGAGGCG  CTTACGCTC  AGCTCTCT  AAGAGAGG  CTTGCGCGG
901  CTGACAGAG  CTTAGCTT  GACTCTGAG  ATGTTGGCG  TGGGCCCTAA
951  GGGGAGGAG  AGCATGGCG  CCGCTGTCT  CATCTGAAC  CAGTATGGG
1001  AGTGGTTT  TGACAGTAC  CTCAACCA  CTAGCGCTT  GAGGACTAT
1051  AGACAGCGG  TACGTGGAT  TCGGCTGAG  AACTCAGC  AGGGAGAGA
1101  GCTTGAAGT  GTTCAGAGG  AAGTGGAGA  GATGCTGAG  GGCAGATTC
1151  TATGGGGCA  CCGCTCTAT  AATGACTAT  AGGTACTTT  TCTTATCAT
1201  CCAAAAAGA  AGATTGGGA  CACACAGAA  TATAACCTT  TCAAGATCA
1251  AGTAAAGAT  GGAAGCGCT  CTCTGAGCT  ACTTTCAG  AAGATCTTG
1301  GCGCTCAGT  CCAGAGGCG  GAGCACTGT  CAATCAGGA  TGGCAGGCA
1351  GCAATGAGG  TGTAGTCA  GTGAGAGAG  GAGTGGGGA  GCATGGCGG
1401  AGACAGCGG  CCGCTCTCT  CTCTCCAGA  CCACTGCA  GACGAGCGT
1451  AGCATCTCT  CCGTCTCT  GCTGCGGCC  CCGTACAG  GCATGTGAC
1501  CAGTACAG  GAGATCAG  ATCTCCGAG  ACTGCCACT  CTGTGAAC
1551  CTTTTCAGAA  TCATGGAGA  GGGCGTGG  GTGTGCTAC  TGAGAAGTC
1601  CTCTTCTCT  TTGACTTTT  GGTCTGAAC  CTGCTCTAC  TGTCTGTTG
1651  TTTTGGGCC  CGATGTTCA  GGTGGGGAG  CAGGACCG  CATGGGCGG
1701  CTTGCGCAG  CTTTACCG  TGCTGACCG  CTGGGCCAG  TGAGTTGGG
1751  GCTTGTGGG  CGGAGTCCA  TACGCTGT  TCACTGCCA  TCTTGGTGA
1801  CAGCTGGG  TGAGTCTCT  AGCAGCTCT  TCTCGAGAG  CAGATTTTC
1851  CTCCTTTGT  TTAGGGACA  TAACAAGCT  TGTGGGCTT  GAGGAGCCA
1901  GACCGGTGT  CTGCGTCTG  CTCTGAGAG  AGAGTGGCC  GGCACACAG
1951  GTTATCATC  AGGGTTTTC  TGTGGGCTT  TGAATTTCA  GCATTTATC
2001  AATTTAAAT  GGCCAGGT  TGGCTGTGG  GTACCCAGC  AGAGCTTCT
2051  CCGCATAGCA  CAGGATCTG  TTGCGTGGG  AGGCTGACT  CGGTATTTC
2101  TGGAGTGG  CAGATGCG  AACCTTGGG  TATTTAGCT  GGAAGGCTA
2151  TGTGATCTA  GCGGTGCT  TTCTGGCTA  GGCCAGATT  TGAAGCTCC
2201  CTGGAACTA  GAGCCAGGA  CAGCCAGTG  CACTGACAG  GGCACGAGT
2251  CAGAGCTT  ATTGGGCG  CTGAGCTCT  GACAGAAAG  GGGCAGCAC
2301  ACCAGGATG  CGATTTAAA  TAAATGAGA  TTTTACTTG  GAAAAAANA
2351  AAAAAAANA  AAAAAAANA  AAAAAAANA

```

BLAST Results

WO 01/12659

PCT/IB00/01496

Entry NSAC2099 from database EMBL:
*** SEQUENCING IN PROGRESS *** Genomic sequence from Human 9q34; HTGS
phase 1, 2 unordered pieces.
Score = 5053, P = 0.0e+00, identities = 1011/1011
8 exons bp 104219-116190

Medline entries

95157530:
Cloning and expression of a Xenopus gene that prevents mitotic
catastrophe in fission yeast.

Peptide information for frame 1

ORF from 184 bp to 1449 bp; peptide length: 422
Category: strong similarity to known protein

1 MGKARVPASK RAPSSPVAKP GPVETLTREK HXKKKRWKS KAREVSKKPA
51 SGPGAVVRPP KAPEDFSQNW KALQDWLHQ KSOAPEKPLV ISQNGSKKKP
101 KIIOQKKRET SPQVGEENP AGKQQLASPG SVPSGSKMDR KAVPVRTAS
151 GTDHFHGTG KRTNGGIVPE KGIIDHKKRE AKKAAAPPT EEDIMFDWD
201 PADIEAAIGP EAKIARKQL QSEGSVSLV LVKEQAFGL: TRALALDCEN
251 VGVCPGEGES HAARVALVMD YGKCVTDIV KPTPTPTDYR TAVSGIRPEN
301 LKQGEELVY QREYADHAG RLVGHALHW DLVVLFDHP KRIADTQRY
351 KPTKSQVKSQ KPSLRLLSEK ILGLVQQAQ HCSIQDAQAA HRLTVNVKKE
401 WESHADRRAP LLTAPDHCSO DA

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRF2phtes3_7p10, frame 1

No Alert BLASTP hits found

Peptide information for DRF2phtes3_7p10, frame 1

Report for DRF2phtes3_7p10.1

[LENGTH] 422
[MW] 46671.91
[pI] 9.79
[HOMOLOGY] PIK:SS3818 XPM2 protein - African clawed frog 7a-96
[FUNCTION] 03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-42
[FUNCTION] 01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 2e-19
[FUNCTION] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YOL094c] 7a-13
[FUNCTION] 04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S. cerevisiae, YOL094c] 7a-13
[FUNCTION] 99 unclassified proteins [S. cerevisiae, YLR107w] 6e-10
[PROSITE] RGD 1
[PROSITE] MYRISTYL 4
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 6
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[KW] All_Alpha
[KW] LOW_COMPLEXITY 11.37 %

SEQ MGKARVPASKRAPSSPVAKPGPVETLTREKHKKKRWKSKAREVSKKPAKSGPGAVVRPP
SEGXXXXXXXXXXXXXXXXX.....
PRD ccc
SEQ KAPEDFSQNWKALQDWLHKQKSOAPEKPLVISQNGSKKKPKIIOQKKRETSPQVGEENP
SEGXXXXXXXXXXXXX.....
PRD ccc

```

SEG      ACDQDQASRGSVPSGSKHNRAPVPTKASCTENKMGKTATWDPIDPVGRODJEHKKR
SEG      .XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
SEG      PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEG
SEG      AXEAPAPPTTEEDIMFDVDPADIEAIEGPEAKIARQLOGESGSSLSLVEQAOAGGL
SEG      KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
SEG      PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEG
SEG      TRAILDCEHVGVPKGEESMAARVSIWQICEGCTYCTVKTPEPTDYTRATVSGIRFEN
SEG      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEG      PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEG
SEG      LGQSEELVQKVEKXELAGLITLGHVLAJHULVLFVLQPKKIRIDTQYKPKFSOVMSG
SEG      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEG      PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEG
SEG      RPSLAKSLKILGLQVQDQICSLQDGAQGNRLTYWKEWESNARRRRLPTAPHCSD
SEG      chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEG      PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEG
SEG      DA
SEG      .cc
SEG      PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosite for DKFZphes3_7p10.1

P500002	107--111	GLYCOSYLINOLYGLICAC	P00C00001
P500004	107--111	CAMP_PHOSPHO_SITE	P00C00001
P500004	156--160	CAMP_PHOSPHO_SITE	P00C00004
P500003	9--12	PKC_PHOSPHO_SITE	P00C00003
P500005	44--49	PKC_PHOSPHO_SITE	P00C00005
P500003	46--49	PKC_PHOSPHO_SITE	P00C00003
P500003	66--99	PKC_PHOSPHO_SITE	P00C00003
P500005	347--350	PKC_PHOSPHO_SITE	P00C00005
P500003	347--350	PKC_PHOSPHO_SITE	P00C00003
P500003	363--366	PKC_PHOSPHO_SITE	P00C00003
P500005	368--371	PKC_PHOSPHO_SITE	P00C00005
P500006	136--137	CK2_PHOSPHO_SITE	P00C00006
P500006	163--164	CK2_PHOSPHO_SITE	P00C00006
P500006	163--167	CK2_PHOSPHO_SITE	P00C00006
P500006	190--194	CK2_PHOSPHO_SITE	P00C00006
P500006	192--194	CK2_PHOSPHO_SITE	P00C00006
P500006	413--417	CK2_PHOSPHO_SITE	P00C00006
P500007	343--351	TYR_ASPHO_SITE	P00C00007
P500007	342--353	TYR_ASPHO_SITE	P00C00007
P500008	151--157	TRP_STYL	P00C00008
P500008	151--157	TRP_STYL	P00C00008
P500008	221--227	TRP_STYL	P00C00008
P500008	221--227	TRP_STYL	P00C00008
P500016	177--178	RGD	P00C00016

(No Pfam data available for DKFZphtes3_7p10.1)

DNFiphes3_7p9

group: nucleic acid management

DNFiphes3_7p9 encodes a novel 691 amino acid protein with similarity to human nuclear domain 10 protein NDP52.

The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle.

The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819 complex.

The new protein can find application in modulation of viral infections and tumour events.

similarity to nuclear domain 10 protein NDP52

complete cDNA, complete cds, EST hits

Sequenced by BNF2

Locus: /map=329.1 cR from top of Chr12 linkage group*

Insert length: 3003 bp

Poly A stretch at pos. 2957, no polyadenylation signal found

```
1 AAGGTGAGGG GAACAGCTGA TCCGCTGTT GGGAGGACAG ATATCTCAAG
51 GCGAGGATGG AAGAAATCAC ACTAAGCCGG GCACCATCCC GTGGTGAGAT
101 CACCTTTTTC AATGTAGGCG GGACATACAT CCCCACACAC AAGGTGAAAT
151 GTCACTACAC CTTTCCCCCA GGCACCATGC CCAGTGCCAG TGACTGGATT
201 GGCATCTTCA AGGTGGAGGC TGGCTGTGTT CGGATTTACC ACACATTTGT
251 GTGGTCTTCC GTGGCTGAAG GTACACTGTA TGGTTCGCC ATTCACACA
301 GTTCTCAATT CAAAGCAGC TACTTGCCCA AACCCAGGAG TCAGCTCTAC
351 CAGTTCGGAT ATGTGAACCG CCAGGGCCAG GTGTGTGGGC AGAGCCCCCC
401 TTCTCAGTTC GGAGAGCAAA GGGCCATGGA TCAGCTGTG ACCTGAGAG
451 AGGCTTCA CAGGCTCTTG GCACTGTCCA GGCAGAGGCA CAGCGAGCTG
501 TTACAGAACCC AGCTCGATGA GAGCCAGCAA GAACCGAATG ACCTGATGCA
551 GCTGAAGCTA CAGCTGGAGG GACAGGTGAC AGAGCTGAGG ACCGAGTGC
601 AGGAGTCCA CAGGCTCTTG GCACTGTCCA GGCAGAGGCA CAGCGAGCTG
651 ATGGAACAGT ACAAGGGGAT TTCCCGTCCC CATGGGAGCA TCACAGAGA
701 CAGGAGATTC CTGAGCCGGC AACAGGGAGA CCATGTGGCA CGCATCTGG
751 AGCTAGAGGA TGACATCCAG ACCATCAGTG AGAAGTGGT GACGAGGAA
801 GTGAGCTGG ACAGGCTTAG AGACACAGTG AAGGCCCTGA CTCGGGAACA
851 AGCAGAGCTC CTGGGCAAC TGAAGAAGT ACAAGCAGAC AAGAGGAAA
901 GTGAGGCTA GCTTCAGTG GCACACAGGG AGAACATCA CTAAATTTG
951 GACCTGAAGG AGGCGAAGG CTGGCAGAG GAGCAGAGTG CTCAGGCTCA
1001 GCGACTGAAA CACAAGGTGG CCGAGATGAA GGACACCTTA GGCAGGCCC
1051 AGCAGGAGGT GGCCTGCTC GAGCTCTTGA AGACACAGCT TCGAGGCCC
1101 CAGGAGCTTG CAGGCTCAAG CCAGCAGAAA GCCACCTTC TTGGGAGGA
1151 GTTGGCCAGC GCAGCAGCAG CCAGGAGCCG CACCATAGCC GAATACACC
1201 GCGCGCCCT GGAATGGCT GAAATTAAGC CAGAGTGGG TGAGCTGGT
1251 TTGCACTTCA AGGAAGAAA ATCCCAATGG AGCAAGGAGC GGCAGGCGT
1301 GGTGACAGCT GTGAGGCGAG AGAAGGACAA GATCTCGAAG CTGAGTGCAG
1351 AGATACCTTG ATTGGAGAG GCAGTTCAGG AGGAGAGAGC CCAAGACAA
1401 GTTCTCAAGA CTGAGCTGGC CCGGAGAGAG GATCTAGCC TGGTACGTT
1451 GTCAGAAAGT AAGCGGAGC TCACAGAGCT CGGTCAGCC CTGCTGTGG
1501 TCAGAGAGA AAGGAGCAG TTACAGAGAG ACBAACAGGA ATTCTAGAG
1551 TACATGAGA AGCTAGAGC CGGCTGGAG AAGGTGGCAG ATGAGAGTG
1601 GAATGAAGAT GCCACACAG AGATGAGGA GCGGCTGTG GGGCTGAGT
1651 GCGCGGAGC TGTGACAGC TCAGAGAGCG AGTCCACAGA AGCATGAGG
1701 CTCCACCTT ATGGCTTTG TCAGGCTGGA CAGCCAGGCT CTTCTCTTC
1751 TGGGCTTCA GAGGCTTCT CCGTGTGTG CATCAGCAG CCGGCTCCA
1801 TTCTCTCTCA CTCTCTGGG CCAGCTGAGG ACAGTAGCTC TGACTCGAG
1851 GCTGAGATG AGAATCAAT CTTATGCGA CTTCTCAAG GTGGGGGGA
1901 GGAGGCCAAC TTACTCTTC CTGAAGTGG CAGTGCCTC TATGACATG
1951 CCAGTGGCTT TACAGTGGT ACCCTGTGAG AACCCAGCAC TGGGGGCGCT
2001 GCAACCCCA CATGAGAGA GTTCTATC TGTAGGAGC GTTTCTCTC
2051 TGAGAGTAC AAGATGCCC TGGAGACCA CATGATGGA CACTTCTTT
2101 TCAGCACCCA GAGCCCTTC ACCCTTGAGT GATCTTACTC CCTGTACAT
2151 GCACAAATAC ACATCATGC ACACACACAC TCACACATAT GCATACACTT
2201 AGGTTTCATG CCAATTTTCT ATCACACTGG GCTCATGAT ATTCTGTCC
2251 CTAGAACTG CTCTGTGTG CCGTGTCTTC ATCCCAAGAT TTCTCACTTC
2301 ATCTCTCTCT ACCTGCTCT TTGTCTCCAG GGAGGGCTTC TGTGTGGAG
2351 CAGTGGCTGA ATTTATCCCC TGAAGTGGT TTTGGAGGA CCGGATGGA
2401 GGAGGCTTCC CCGTGTGGA ATAGATCGT CCACCTCTAG CCCTGTTGC
```

WO 01/12659

PCT/IB00/01496

```
2451 TTCTGATACA CAGCCACTGC ACACACACAC TCACACTCAC ACTCCCTTGT
2501 CTGATGCCGC AAGGCCAHTT CCTGGGGCAC CTTACCTCTT CTTATTGGGA
2551 GTTTCGGTNG GTTACCTGCA GTTTCCTCTG GGGTCTGCAC AGAGGCAGCA
2601 GCATGGACAT CATGCCCTCT CAGGTCCCTT TTGGTTCTCA GTTTCATTGG
2651 TTCTCTCTTC TGTGCCCA TGACCTCTGC TGGCCACACC TACCTTTTC
2701 CATAACCTTA GGTATTCAGT TTGGAGGGGT TTTTGTATT TTTGAGGATT
2751 CCTGTATTCT GTATCTCTCT CTGCACTCTC CTCACATGGA AAGAATAAAT
2801 GTATTCTGTC CTTCTGTGAG GAATGGGGGG AACACCTCTT CCGAGTATTC
2851 CCAATTTCGA AGGCCCCCTT CCTCTCCAG GTCCCCCACC AGCAATAAAA
2901 GCTTCCCTCT GATATCCATC CCTTTGTAGT TTGAACAAT ATATTATAT
2951 GATATTGAAA AAAAAAAA AAAAAAAA AAAAAAAA
3001 AAA
```

BLAST Results

Entry HS189353 from database EMBL:
human STS W1-11261.
Score = 2191, E = 1.4e-92, identities = 463/485

Medline entries

95310349:
Molecular characterization of NDP52, a novel protein of the
nuclear domain 10, which is redistributed upon virus
infection and interferon treatment.

97378672:
Cellular localization, expression, and structure of the nuclear
dot protein 52.

Peptide information for frame 3

ORF from 57 bp to 2129 bp; peptide length: 691
Category: similarity to known protein
Prosite motifs: RND (557-560)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (475-497)
LEUCINE_ZIPPER (482-504)

```
1 MEESPLSRAP SRGGVFLNV ARTYINTEV ECVTLPGOT MPASDMIGI
51 FEVEAACVAD VITVYMSVSP ESTTDCSPIN TSVTOASTL PKPGALYOF
101 RYVNRQGVQC GQSPFFQFEK FRPHDELVTI EADGGSDIL LVVPKATVLO
151 MQLDESQQR NDINLKLQI EGQVTELRK VQLEENALAT ARGENTELNE
201 QKGIIRSHQ EITREKILS RQGGQVARI LELEDITCTI SPVLTIEVE
251 LDRLRDTVKA LTRZEKLLG QIAEQVQDE OSEAEIQVQ QENHHLWLOL
301 KEARSMQEDQ SAQGRLEKQ VQQRDTLQD AQQRVALEP LKQLKQAOE
351 LAASSQRAY LQELASASA AARQRTIAGI RREPLVYAV NQRLALGLW
401 LKEKCOMSK ERAGLQSV EAKDKILLEL AEILRLKAV OEERTOMOVF
451 RTELARKQS SLVQLSEAR ELTELASALR VLQKREQLQ EKKQELIEM
501 RLELRLEKVP ADEMNQNT TEDEKAVGL SPALATTEE DESPEOMHLP
551 PYGLCEKGDG GSSPAGPREA SPLVVISQPA PISPHLSGPA EDSSSDSEAE
601 DEKSVIAAV OSQGERANL LPELGSAPTD MGGITTVGTL SETSTGGPAT
651 PTHKECFICK EAPFASDQD ALZDNDGCHF FTSTQDPPTF E
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZptea3_7p9, frame 3

PIR:A56733 nuclear domain 10 protein NDP52 - human, N = 2, Score = 307,
P = 7.7e-28

TREMBL:AB008852.1 gene: "NDP"; product: "NDP52"; Bos taurus mRNA for
NDP52, complete cds., N = 2, Score = 302, P = 4e-27

TREMBL:AC004549.1 gene: "NDG5C.H NC459W13.1"; product: "TXP151"; Homo
sapiens BAC clone NC459W13 from 7p15, complete sequence., N = 2, Score
= 275, P = 2.3e-25

PIR:G02043 TXBP151 - human, M = 2, Score = 270, E = 0.5e-25

TREMBL:DW35816.4 gene: "zip"; product: "nonmuscle myosin-II heavy chain"; Drosophila melanogaster nonmuscle myosin-II heavy chain (zip) gene, complete cds., M = 1, Score = 254, E = 1.4e-17

>PIR:AS6733 nuclear domain 10 protein NDP52 - human
Length = 446

HSPs:

Score = 307 (46.1 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
Identities = 104/323 (32%), Positives = 158/323 (48%)

Query: 15 VNFLVARTYIPNTEVECHYTLRPGTHPSASDMIGIFKVEACVROVHTFVMSVPESTT 74
V F +V + YIP V CHYT +P DMIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFMSVEKFTYIPGGDVTCYHTTQNFIPRAKDMIGIFRVGKTKTRETITTFNVTLPIDLN 82
Query: 75 DGSPIHTSVQFQASTLPKPGALYQFTYVNRQGVCGSPFPQFREPRLDELVTLEAD 134
+ S VQF+A YLPR + YQF IV+ G V G S PPOFR D LV +
Sbjct: 83 HSAKQGVQFQATYLPDD-ETTYQFTYVDEGVVAGASIPQFRPEHEEDILVVTQ-- 139
Query: 135 GGSILLVVPKATVLOH-LDES--QDENLMLQLKQLQGVTE-LASRVQELERALA 189
G + + K +HQ L +S Q+H HQ +LQ + + S L+S +LLE +
Sbjct: 140 GEVEEIEQHNKELCKENQELKDSGISLQKQNSDMQELQKQEELETQSNKKLELKV 199
Query: 190 TARQ-HTELMEQYKGISRSNGEITEEDI--LSRQGDHVARILELEDDIOTISEKVLTK 247
+ TEL+ Q K ++ E+ I + + Q + E+E +Q +E T+
Sbjct: 200 EQDYMETELL-QLKEQMKSSSEKNGIRVDQLQGLSTQEKEMKLVGGDQK--TE 256
Query: 248 EVE-LDRLKDTVKALTRQEKLLQGLKEVQADKEQSEALQVQDENHMLDLKAKSW 306
+E L + D + EQ K +L+ +QE OOE H DL + S
Sbjct: 257 QLEQLKENDHFLSLTEQRKDKKLEQTVQKQKQNTTAMKQQLDENFDLSTKLSE 316
Query: 307 QCEQSAQQLKADRVQKHTLQGAQQRV 315
E OR +++++ D L + R+
Sbjct: 317 NEIICHALQKQERLEGENDLERENSRL 345

Score = 304 (45.6 bits), Expect = 2.1e-27, Sum P(2) = 2.1e-27
Identities = 98/337 (29%), Positives = 163/337 (48%)

Query: 15 VNFLVARTYIPNTEVECHYTLRPGTHPSASDMIGIFKVEACVROVHTFVMSVPESTT 74
V F +V + YIP V CHYT +P DMIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFMSVEKFTYIPGGDVTCYHTTQNFIPRAKDMIGIFRVGKTKTRETITTFNVTLPIDLN 82
Query: 75 DGSPIHTSVQFQASTLPKPGALYQFTYVNRQGVCGSPFPQFREPRLDELVTLEAD 134
+ S VQF+A YLPR + YQF IV+ G V G S PPOFR +E
Sbjct: 83 HSAKQGVQFQATYLPDD-ETTYQFTYVDEGVVAGASIPQFR--PENE----- 130
Query: 135 GGSILLVVPKATVLOHQLDESQDERNDMLQLKQLQGVTELASRVQELERALATARQ 194
DILV Q +++E +Q +L+ +L+ L+ + +L +QE
Sbjct: 131 --EDILVVT-----QGEVEEIEQHNKELCKENQELKDSGISLQKQNSDMQELQKQEELE 182
Query: 195 HTLMEQYKGISRSNGEITEEDI--LSRQGDH-VARILEDDIOTISEKVLTRVELDR 253
E ++ I ++ ++ ++Q D+ +L+L+ Q +S + +D+
Sbjct: 183 ELLETQ-----INRLELKVQKQDWTETELQLKQKQKSSSEKNGIRVDQ 232
Query: 254 LRDYKALTRQEKLL--QLKEVQAD--KEQSEALQVQDENHMLDLKAKSWQ 308
L+ + +E EKL- Q R Q + ES L + Q L+ + Q
Sbjct: 233 LQALSTQEKEMKLVGGDQKQNTTAMKQQLDENFDLSTKLSE 316
Query: 309 EQSA--QADLKADRVQKHTLQGAQQRV 315
E +A + Q L D+ + L + + L+ EE+L G +L
Sbjct: 293 EYTMKQQLDENFDLSTKLSENEIICHALQKQERLEGENDL 337

Score = 124 (18.6 bits), Expect = 2.3e-06, Sum P(2) = 2.3e-06
Identities = 53/227 (23%), Positives = 113/227 (49%)

Query: 138 DILLVVPKATVLOHQLDESQDERNDMLQLKQLQGVTELASRVQELERALATARDITE 197
DIL+V Q +++E +Q +L+ +L+ L+ + +L +QE E
Sbjct: 132 DILVVT-----QGEVEEIEQHNKELCKENQELKDSGISLQKQNSDMQELQKQEELE 185
Query: 198 LMEQYKISRSNGEITEEDI--LSRQGDH-VARILEDDIOTISEKVLTRVELDR 256
++ I ++ ++ ++Q D+ +L+L+ Q +S + +D+L+
Sbjct: 186 TLQS-----INRLELKVQKQDWTETELQLKQKQKSSSEKNGIRVDQLQ 235
Query: 257 TVKALTRQEKLLQGLKEVQADKEQSEALQVQDENHMLDLKAKSWQEQSAQQR 316
+ +E EKL VQ D++++E +L+ ++ER HL L L E + O++ ++
Sbjct: 236 QLSTQEKEMKLVGGDQKQNTTAMKQQLDENFDLSTKLSE 316

[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YBR289w] 4e-06
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YBR289w] 4e-06
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YML250w] 4e-06
[FUNCAT] 03.13 meiosis [S. cerevisiae, YML250w] 4e-06
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-05
[FUNCAT] 38 classification not yet clear-cut [S. cerevisiae, YJR134c] 4e-05
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-05
[FUNCAT] 08.19 cellular import [S. cerevisiae, YML243w] 7e-05
[FUNCAT] 01.02.16 polynucleotide degradation [S. cerevisiae, YML243w] 7e-05
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YML243w] 7e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YML079c] 2e-04
[FUNCAT] 03.01 cell growth [S. cerevisiae, YML079c] 2e-04
[BLOCKS] BL006828 2P domain proteins
[EC] 3.6.1.32 myosin ATPase 1e-13
[PIRKEW] nucleus 6e-10
[PIRKEW] phosphotransferase 2e-07
[PIRKEW] duplication 1e-07
[PIRKEW] citrulline 1e-09
[PIRKEW] tandem repeat 1e-13
[PIRKEW] heart 5e-11
[PIRKEW] endocytosis 5e-09
[PIRKEW] polymorphism 3e-06
[PIRKEW] cornified cell envelope 1e-06
[PIRKEW] transmembrane protein 6e-12
[PIRKEW] serine/threonine-specific protein kinase 2e-07
[PIRKEW] cell wall 1e-06
[PIRKEW] zinc finger 3e-09
[PIRKEW] metal binding 3e-09
[PIRKEW] DNA binding 4e-08
[PIRKEW] muscle contraction 1e-11
[PIRKEW] IgG constant region-binding 1e-06
[PIRKEW] acetylated amino acid 4e-09
[PIRKEW] actin binding 1e-13
[PIRKEW] mitosis 9e-05
[PIRKEW] microtubule binding 9e-09
[PIRKEW] ATP 1e-13
[PIRKEW] chick filament 1e-10
[PIRKEW] phosphoprotein 1e-13
[PIRKEW] epidermis 1e-06
[PIRKEW] leucine zipper 1e-07
[PIRKEW] glycoprotein 4e-07
[PIRKEW] skeletal muscle 4e-10
[PIRKEW] disulfide bond 1e-07
[PIRKEW] calcium binding 1e-09
[PIRKEW] alternative splicing 1e-10
[PIRKEW] coiled coil 1e-13
[PIRKEW] P-loop 1e-13
[PIRKEW] heptad repeat 6e-10
[PIRKEW] methylated amino acid 1e-13
[PIRKEW] basement membrane 3e-06
[PIRKEW] immunoglobulin receptor 2e-07
[PIRKEW] peripheral membrane protein 5e-09
[PIRKEW] dimer 1e-07
[PIRKEW] cardiac muscle 1e-10
[PIRKEW] extracellular matrix 3e-06
[PIRKEW] hydrolase 1e-13
[PIRKEW] microtubule 6e-10
[PIRKEW] muscle 2e-09
[PIRKEW] membrane protein 3e-06
[PIRKEW] EF hand 1e-09
[PIRKEW] cytoskeleton 6e-12
[PIRKEW] hair 1e-09
[PIRKEW] calmodulin binding 5e-09
[PIRKEW] Golgi apparatus 3e-09
[SUPTAM] myosin heavy chain 1e-13
[SUPTAM] conserved hypothetical P115 protein 1e-08
[SUPTAM] hypothetical protein K1074c 5e-07
[SUPTAM] centromere protein E 9e-09
[SUPTAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-07
[SUPTAM] calmodulin repeat homology 1e-09
[SUPTAM] myosin motor domain homology 1e-13
[SUPTAM] alpha-actinin actin-binding domain homology 3e-13
[SUPTAM] tropomyosin 3e-07
[SUPTAM] plectin 3e-13
[SUPTAM] trichomyosin 1e-09
[SUPTAM] plectatin repeat homology 4e-06
[SUPTAM] ribosomal protein S10 homology 3e-13


```
(SUPFAM)          gliadin 3a-08
(SUPFAM)          protein kinase homology 2e-07
(SUPFAM)          protein kinase C zinc-binding repeat homology 4e-06
(SUPFAM)          involucrin 1e-06
(SUPFAM)          kinesin motor domain homology 9e-09
(SUPFAM)          human early endosome antigen 1 5e-09
(SUPFAM)          unassigned kinesin-related proteins 8e-08
(SUPFAM)          M5 protein 3e-08
(SUPFAM)          cytoskeletal keratin 3e-08
(PROSITE)         LEUCINE ZIPPER 3
(PROSITE)         RGD_1
(PROSITE)         MYRISTYL_6
(PROSITE)         CK2_PHOSPHO_SITE_25
(PROSITE)         PKC_PHOSPHO_SITE_6
(RM)              All_Alpha_9.12
(RM)              LOW_COMPLEXITY_39.36
(RM)              COILED_COIL_39.36

SEQ      MEKSLKAPSRGCVNLAIVARTIIPMTVECHYLPQTHPSASDWIGIPYEAACVMD
SEG      .....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
COILS    .....

SEQ      YHTFWSSVPESTTDSPIHTSVQFQASYLPKPAQLVQFRVYWRGQVCCQSPFPQFRE
SEG      .....
PRD      #####cccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

SEQ      PAPHDELVTLEADGSDILVVPKATVLMQLDESQQRMDLMQLQLGEGVTELRSR
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

SEQ      VOELERALATARQENLEADYEGISNMGITEARDLSRQQQDHWARIILEDDGIOTI
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      SEKVLTKVELDLRLDTVKALTRDEKLLGQLEVPQADKQSEAELOVAQDHNLKLOL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    ..... CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      KEAKSWQEQSAQAQRLKDKVAQKDTLGQAQQRVAELEPLKEQLAGAQELAASSQKAT
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      LIGELASAAAAADRTIAELHRSLEVAEVMGLAEGLHLKEEKQCMKERRAGLIQSV
SEG      .....
PRD      #####XXXXXXXXXXXXX.....
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      AERDKILKLSAEILALEKAVQERTQWQVYKTELAREKDSLVOLAESKRELTELASALA
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      VLQKERQEQKEQELLEYNLEALEKVADEKVMEDATTEDEEAAGVGLSCPAALTOSE
SEG      .....
PRD      #####XXXXXXXXXXXXX.....
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      DESPEOMRLPPTGLCEAGDPSSFPAGPREASPLVVISQPAFISPHLSGPAEDSSDSTAE
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....

SEQ      DEKSVLMAVQSOGGEAMLLPELGSFTDKASGFTVGLSETSTGGPATPTWKECPICK
SEG      .....
PRD      #####XXXXXXXXXXXXX.....
COILS    .....

SEQ      EATPAESDKDALEDHNGHFFFTQDPTTIE
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
```

WO 01/12659

PCT/IB00/01496

Prosite for DKFZphtes3_Tp9.3

PS00005	190->193	PKC_PHOSPHO_SITE	PD0C00005
PS00005	241->244	PKC_PHOSPHO_SITE	PD0C00005
PS00005	257->260	PKC_PHOSPHO_SITE	PD0C00005
PS00005	468->471	PKC_PHOSPHO_SITE	PD0C00005
PS00005	652->655	PKC_PHOSPHO_SITE	PD0C00005
PS00005	667->670	PKC_PHOSPHO_SITE	PD0C00005
PS00006	28->32	CK2_PHOSPHO_SITE	PD0C00006
PS00006	43->47	CK2_PHOSPHO_SITE	PD0C00006
PS00006	68->72	CK2_PHOSPHO_SITE	PD0C00006
PS00006	72->76	CK2_PHOSPHO_SITE	PD0C00006
PS00006	129->133	CK2_PHOSPHO_SITE	PD0C00006
PS00006	156->160	CK2_PHOSPHO_SITE	PD0C00006
PS00006	208->212	CK2_PHOSPHO_SITE	PD0C00006
PS00006	239->243	CK2_PHOSPHO_SITE	PD0C00006
PS00006	282->286	CK2_PHOSPHO_SITE	PD0C00006
PS00006	305->309	CK2_PHOSPHO_SITE	PD0C00006
PS00006	376->380	CK2_PHOSPHO_SITE	PD0C00006
PS00006	383->387	CK2_PHOSPHO_SITE	PD0C00006
PS00006	468->472	CK2_PHOSPHO_SITE	PD0C00006
PS00006	520->524	CK2_PHOSPHO_SITE	PD0C00006
PS00006	537->541	CK2_PHOSPHO_SITE	PD0C00006
PS00006	539->543	CK2_PHOSPHO_SITE	PD0C00006
PS00006	543->547	CK2_PHOSPHO_SITE	PD0C00006
PS00006	593->597	CK2_PHOSPHO_SITE	PD0C00006
PS00006	595->599	CK2_PHOSPHO_SITE	PD0C00006
PS00006	597->601	CK2_PHOSPHO_SITE	PD0C00006
PS00006	612->616	CK2_PHOSPHO_SITE	PD0C00006
PS00006	639->643	CK2_PHOSPHO_SITE	PD0C00006
PS00006	652->656	CK2_PHOSPHO_SITE	PD0C00006
PS00006	667->671	CK2_PHOSPHO_SITE	PD0C00006
PS00006	683->687	CK2_PHOSPHO_SITE	PD0C00006
PS00008	39->45	MYRISTYL	PD0C00008
PS00008	107->113	MYRISTYL	PD0C00008
PS00008	204->210	MYRISTYL	PD0C00008
PS00008	414->420	MYRISTYL	PD0C00008
PS00008	561->567	MYRISTYL	PD0C00008
PS00008	613->619	MYRISTYL	PD0C00008
PS00016	557->560	RGD	PD0C00016
PS00029	163->185	LEUCINE_ZIPPER	PD0C00029
PS00029	475->497	LEUCINE_ZIPPER	PD0C00029
PS00029	482->504	LEUCINE_ZIPPER	PD0C00029

(No Pfam data available for DKFZphtes3_Tp9.3)

DKF2phtea3.8e24

group: signal transduction

DKF2phtea3.8e24.3 encodes a novel 658 amino acid putative GTP-binding protein, related to yeast YGL099w and mouse MOR1 putative GTP-binding proteins.

GTP-binding proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

strong similarity to guanine nucleotide binding proteins

complete cDNA, complete cds, potential start at bp 31, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 3290 bp

Poly A stretch at pos. 3269, polyadenylation signal at pos. 3251

```
1  CGTCCAGCGG TCGTGTGCG ATGGGCGGGA GGAGAGCCGC GGCCTGTGG
51  TCGCTGGGAC GGGGCTTTAT GCGCCATGAG ACTGAGCGGA GCGAAGACCA
101  TCTCTACACT GACTCTCTGT TCGACACAGG TGACTCTAAT GATGGCTATG
151  ATTGGGGTGG TCTTAATCTT CAGTCAGTGA CTGAACAGAG CTCCTTGAT
201  GACTTGGCTT CTACTGACAG ACTTCAGAGA ACAGACTTTG TACTGTAAAA
251  ACTTAATATT AACTTGTGCG CTGCTGAGGC TAGAATCGGA CTACTGTCTT
301  TCGAGGAGAG CCGAGAAATT AAGAAGCTCC ATGAAGAAAA CAACAGCTTC
351  TTGTGTATAG CGAGGAGAGC AACTGCAAC CAATATACTA CCGACAGAGA
401  ACTCAACAAA CGAGAGAAAG ATAATCTTCT AGAATGAGGA GGTACAGCTG
451  TCCGGCTAGA AGAGGACACG AAGCTGATAT TGACTCCATT TGAACDAAAT
501  TTGACTTTT GGGCGGAGCT CTGGAGAGTC ATTGAGAGAA GTATATTCT
551  GGTCCAGATA GTAGATGCTC GAACCCCACT CCTGTTTGA TGTAGGATT
601  TGAATGTGTA TGTGAAGAA ATGATAGCCA ATAGGAGAGA GTTATTCTG
651  ATCAACAGCG CAGACTTGCT GACTGCTGAG CAGCGGAGTC CCTGGGCTAT
701  GTACTTCGAA AAGAAGATG TCAAGGTTAT TTCTGTGCTA GCTTGGGCGG
751  GAGCATTGCG CTGTAATGCT GACTCTGAGC AAGGGGCA CAGAGATGAT
801  AGACAAAGCA ACACAACTGA GTTTGGACAT TCCAGTTTCC ACCAGGCTGA
851  AATTTCGCAC AGTCATCCCG AACATCTCCG AGTAGGGAT TCTCTTCC
901  TTAGTAAAAA TCCACAGACG GATGAAGATC AACTGATGTT TCAGGACTGT
951  CGAGAGGAGG AGGAAGACGA CTGGCAGAGC TGCTCAGAAI AAGACGGTCC
1001  CAGGAAGAGG GACTGCAGCC AGGACTGCGA GGAAGCTCTC ACTCGAAT
1051  CTGGCTGCTG GAGAGGAAAC ACCCCAGAGA AAGGGKAGAT ACACAAATTT
1101  AGCCATCTGG TATCCAGACA GGAGTTACTG GAGCTCTTTA AGGAGCTACA
1151  CACTGGGAGA AAGCTGAAGC ATGGGCACTC TACGCTGGGA CTGTGGGCT
1201  ACCCTAATGT TGGTAAGACT TCACAACTCA ACACCATCAT GGGCAGCAG
1251  AAGATATCTG TGTCTGCCAC ACCTGCTCAC ACAAGGCACT TTCAGACTCT
1301  CTATGTGGAG CTGSGGCTCT GCCTGTGGTA CTCTCTCTGC TTGCTGATG
1351  CATCTTTTGT GTCTACCAAG GCAGAAATGA CTGGCAGCGG AATCTCTCCA
1401  ATTGATCAGA TGAGAGATCA TTTTCTCTCT GTATCACTAG TTGTCAGAA
1451  TATTCCAGCA CATCTTTTAC AAGCTACTTA TGGCATTAGC ATCAATAGCC
1501  CTAGAGAGGA TGAAGATCCC CACCGACTTC CAACATCGGA AGAATCTTTG
1551  ACAGCTTATG GATACATGCG AGATCTATG ACAGGCTATG GACAGCGAGA
1601  CAGCTCTGCG TCTGCGGCTC ACATCTTGAA GGACTATGTC AGTGTAAAGC
1651  TGTCTGACTG CATCTCTCTC CTGGAAGAG ATCCTGTATC TTTTCAGCAT
1701  CAGACACAGC GACTCTTAGA GAAATAAATC AACAGTATC AATTAATAT
1751  GCAGCTAGCG AGAATATAAA AAGCAAGCA GATTGAATAT ATCTTGACA
1801  AAGCTTTTTT GCATCAAGAG AATGTAGGGG CTTTGACCAA AGGATCTCAG
1851  CTTGTATGAG GTTACAGCTC CGGAGCTGCT GTACTGACTC CATCTACTG
1901  GAGCTCTGAG AACGGGGGCG GGAAGGCTCT GAAAAACAT GGCACAGAAA
1951  ATMAAAGAGA AAAAAGTCTG AGACTCTACA AGCCTGGA TATGTAGGT
2001  TGGGTTCGCA CAGAAATGTT ATTCTGATTC TGCAGATGGA AAGAGCAGA
2051  AGCTGGCTGT TGGCTGTGGA ACTGTCCGCA GACACTAGA CTTAGAACG
2101  GGCTCTGCTC TTGAGAGCA CGGCTGACAC CAGACTCTC CATCTAGAGA
2151  CCAAGGGGCT CCTGGAACCA CCGACTCTGA CAAAAGGAG TCATCTGGA
2201  GCGCGAGAA CTACTCTCTG GCGCGGCACT GTGGCTCAGC CACCAACATG
2251  GAGAAAGGCG GTCTCTACTA AATATACAAA AATATTAGC AGGCTGGTG
2301  GCGCGAGACT GTAACTCCAG CTACTCTGGA GCGTGAGGCA GGAATATCAC
2351  TTGAACGAGG GAGGACAGGT TTGAGTGA TGGAGATTGC GGCGCTGCMC
2401  TCAACCTGCG GGCACAGCTC GAGACTGATC CACAGAAAA AATTTTCCA
2451  AGGATGCTT CACGACACAC ATTGGGAGC AGCTGAAG AGAATTTCC
2501  CATCTGAGT GTCTGATGTC GCTTCTCCG ACTTAACA AGGACTTGG
2551  CTCACTTAG TCTACAGCG GGGCTACACT CCACTAAGA ACATGTAGAA
2601  TGTACACAC GGTGACAGG GAAGCTGCGG TATTACTAC CTAGCCCCCA
```

WO 01/12659

PCT/IB00/01496

```
2651 TCTTCACYGG TTATTCACCT TATTAAAAAT GTCCAGATA AGCAATCTC
2701 CATATAGAGG AAGTAGATTA GTGTTGCTT CGGATGGGA GGAATGGGA
2751 GATTGAGTC TTCTTTTGG AGTATAAA ATGCTCTAAA ATTGACTGTA
2801 GCGATGCTCA CACAACCTG AATATGCTTA AGACCATGA ATTACACACT
2851 TTACGTTGCT GAATTGATG GTATGTAAAT TATAGTTCAA TACATAGTT
2901 ACAAAAGATA ATCAAAAGCA TGAAGGACT ATTGATGTGG TTGATCTG
2951 TGTCTCACC GAGTCTCATG TTGAAATGTA AGCCCTCTGG TGGAGGCGA
3001 TGGGATTATG GGGCAGAGTC CTCACAAGC GTTAGACAC ACCGCTCAG
3051 TGCTGCTTCT CTGATATCTT CTTCTCTCTA CATCTGCTTG CTGAAAGTC
3101 TGTGGGCTCT CCCTCTCTCT TCCCTCTCTG TCTGGGCATA TAAGATGTC
3151 CTGCTCTCTG TTGCGCTCTG ACATGATGTC TAGTCTTCTT GAGGCTCTCT
3201 TACAGCAAAA AGCTGCTCTG CTCTCTCTAG CATCTACTGG ACCGTGAGCC
3251 AATTAAAGCT CTTTCTTTTA TAAAAAAGA AAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 21 bp to 1994 bp: peptide length: 858
Category: strong similarity to known protein

```
1 MGRARAPAG SLGALAKHO TORSRSKHT DSWLMTSELN DQYDNGRLML
51 QSVTQSSLD DFLATLALG TEFVAKLMI KPVPAKATG LLSFRESQMI
101 KLMSEKPOF LCIPRPMHM QTTTPEELD AKQWFLDM RQVMALEED
151 KLILTFPERM LQPRQLMNV IESDVIQVI VDAHMLLP CEDLECVRK
201 WDAKENVIL INKADLTAE QRSAAHYTE EDQVIVPS ALAGAPLNG
251 GSEEEARRDD QSHHTTEFGH SSFOQALSH SESELPARD SPSEINPTT
301 QDDSEYKNC PEEZEDDMYT CSEEDCFKEE DCSQOMKSS TADSEARSK
351 TPQRQLNHH SILVSPQELL ELFEKELTOR VFEQDQTYG LVGVNPKES
401 STINTINGNA KVSVSATFGH TRNPOTLYVE PGLCLDCPG LVHSPVSTK
451 ADHTCSGILP IDQKRDVFP VSLVQCMFM VYLEATYGIH IYFREDKDP
501 HSPFTSEILL TAYOTMGMN TANGCTQPM SARTLEQDY SELLICHPH
551 PGROPPTFGH QHRLLEKHM WSEIEMQLG RKKAKOIEH IVDRTFFHQE
601 HYKALTKGV AVHGTPOSG VYTASTASSE WGAQWPKKH GHRKREKSR
651 KUYELDM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFiptes3_Be24, frame 3

SWISSPROT:YANG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME 1., N = 3, Score = 560, P = 1.6e-111

PIR:S64106 hypothetical protein YGL099w - yeast (Saccharomyces cerevisiae), N = 2, Score = 344, P = 2.6e-105

TRINHL:CEAF3143.1 gene: "C53H9.2"; Caenorhabditis elegans cosmid C53H9., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:NMRL_MOUSE POSSIBLE GTP-BINDING PROTEIN NMRL., N = 2, Score = 311, P = 7.5e-31

>SWISSPROT:YANG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME 1.
Length = 616

KSPs:

Score = 560 (84.0 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 119/253 (47%), Positives = 162/253 (64%)

Query: 12 LGRAAKHQTORSRSKHTDSWLMTSELNLDQYDNGRLMLQSVTQSSLDQFLATLALG 71
LGRA+ T+ R+ + H+ + R L+SVT ++ LD+FL TAEI

```

Sbjct: 12 LGRAIOSDTKNNRKK--GGLKNIIVDSDFK--KAALRSVTHETDDEFMTAEIGEV 67
Query: 72 EFVAEKLAIKFVP--AARTGLSEFEKQRIKLIHEHGFOLIPRRPNHMQWTFPEELKQ 130
EF+AEK M+ + E LLS EE+ R E+ E+HK L IPRRP+H+Q TT EL +
Sbjct: 69 EFIAEKQNTVIONFEQHPFLSKKEEAARSKQKQKRNKRLTIPRRPNHMQWTTAVELDR 127
Query: 131 AEKDFLEWRRQLVLEEEKRLITFFERKLDPMQLWVIERSDIVVQIVDARUPLIFR 190
E++FL NRR L -L++ + I-TFFERKL+ WQLWVIERSD+VQIVDARUPL FR
Sbjct: 128 HEKESFLERKDLACQDVEGCIIVPTFERKLEIHWGLWVILKSDIVVQIVDARUPLIFR 187
Query: 191 CZGLQYVVEDDANZHVILLHKAADLTADQSAHANYFEKDWKVIWSEALAGLPLMG 250
LE TVKE+ +RN +L+KAD-LT EQR+ M+ YF + + +FISA A H
Sbjct: 188 SAHLEQTVKEVQPSKFNLLVHKAADLTEDQRYHSSYFVNDNIPLFFSANMAA-EAKE 246
Query: 251 DSEEEKARRDQDSH 264
E+ + SM
Sbjct: 247 RGEDETVESTSSM 260
Score = 532 (79.8 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identifies = 131/323 (40%), Positives = 192/323 (59%)
Query: 340 STADSEARSRTKPNQIHWFSNLSKELLEFEKLTGRVKGQ--LTVGLVGMV 397
ST+ +E + +H+ S + + + L +F+ + + + DG+ +T GLVGMV
Sbjct: 256 STSNEIPELQADENHVS-SRIATLRVLEGIFEFAS--YLPQGTMTITVLVGMV 312
Query: 398 GKSSTINTINGKKSVSATPQTKHPTLYVEGCLCDQPLVHPSFVSKADHTCSG 457
GKSSTIN +G+KKSVS+TNG TKHPT+ + + L DQCLV PSF +T+A+ + G
Sbjct: 313 GKSSTINALGKSVSVSTGTHPTUTLBLEKVELLQCLVPSFATYQADVLQDG 372
Query: 458 LPIIDQNDHVPFVLVQVIFWPLVATYGIW-ITPREDPMPPTSEELLTATYCN 516
-LPIIDH++ P +L+ + IP+ VLE Y I I P E E P+++E+L +
Sbjct: 373 VLPIDQLREYTPSALMAERIPEVLETLTIRIRIKPIE-EGGTGVPSAQVLPFARS 431
Query: 517 RGFMTAH--GPDQPSARTILEDIVSGKLLTCHPFPQ--RDPYTFQHOHQLLEHRRSD 573
RGFM AH G PD R+AR -LKDIV+KLLT HPFP F +H + + + SD
Sbjct: 432 RGTWRAHGTGDSAAARILLADYVWGLLVHPPFVHSGSEFEDHGLVSA-TSD 490
Query: 574 EIKHQLGR---HKKAKQIDH-IVKTFTHGH--VVALTEQDAVH-G--TEPSSGVVTA 624
I +L R + + E+ +VD +F QDH VH + RG R G YK + +
Sbjct: 491 BITEKLQRTAISDHTLSAESQVDDQTF-QENPHVAVHVGTAVANOGPVYEGKNTHPFF 549
Query: 625 STASSENAGK-PWKNKGNKKEKSRRL 652
+++ + K P G + E+R+L
Sbjct: 550 ORALKODASFKTPHNGQRFLERRKARQL 578
Score = 47 (7.1 bits), Expect = 1.3e-60, Sum P(3) = 1.3e-60
Identifies = 21/84 (25%), Positives = 35/84 (41%)
Query: 552 GRPVTYQHQRLLLENHNSDEIKHQLGKAKGIEHIVKTFTHGHVVALTEGVGA 611
G D T++ + + +DE + R K+E I +K F TK
Sbjct: 248 GEDLETYSTSMCIPELQADENHVSRIATLRVLEGIFER--FASTLPDGTMTITFG 305
Query: 612 VNGYKPGSGVVYTAASSENAGK 635
+GY P G +ST ++ G+ E
Sbjct: 306 LWCY-PMYC--KSTIALVQSEK 326
Score = 43 (6.5 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identifies = 7/13 (53%), Positives = 9/13 (69%)
Query: 638 KKHGNKKEKSR 650
KKH +NK+ K R
Sbjct: 596 KKHKKKRSEQR 608

```

Pendant information for DNFiphtes3_8e24.3

Report for DNFiphtes3_8e24.3

```

[LENGTH] 658
[MW] 75226.58
[pI] 5.86
[MOLEC] SWISSPROT:YANG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.14C IN CHROMOSOME
I. 5e-56
[FMCAT] 99 unclassified proteins [S. cerevisiae, YC1899w] 3e-55
[FMCAT] + general function prediction [M. larnaschii, M21464] 1e-16
[FMCAT] 08.16 extracellular transport [S. cerevisiae, YER006w] 3e-09
[PIKIM] P-loop 1e-27
[PIKIM] GTP binding 1e-27
[SUPFAM] conserved hypothetical protein MG442 7e-08

```

WO 01/12659

PCT/IB00/01496

```

(POSITE)  ATP_GTP_A      1
(POSITE)  MYRISTYL      3
(POSITE)  AMINATION      2
(POSITE)  CAMP_PHOSPHO_SITE  1
(POSITE)  CK2_PHOSPHO_SITE  19
(POSITE)  TTP_PHOSPHO_SITE  2
(POSITE)  PKC_PHOSPHO_SITE  10
(POSITE)  ASN_GLYCOSTYLATION  2
(EN)      Alpha_His
(EN)      LOW_COMPLEXITY  4.56 %

```

```

SEQ  MGRRAAPAGGSLGAALNRHQTGRSRKRTDSMLNLTSELNDGYWGRALWLGVSVEQSLLD
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  DFLATAELACTEPFAERLWIFVFAEARTLLSTESQR I KKLHEENKOFLCI PRAPRW
SEG  .....
PRD  hhhhhhhhhhecccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  QMTTPEELKQAEONFLWRQLVLEEQKLIITPFEANLDFWQLWVIERSDIVVQI
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  VDAENPLFRCEOLECYVEDNAKEMVILINKADLTAEQSAWAMYFEXEDVKVIFWS
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  ALAGAIPINGDSEZANRDCQSMTEFGHSFDOAEISHSEHLPAKDSPLSDMPTT
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  DEODSEYDCPEEEEDQWTCSEEDGPREEDCSQWGESSTADSEARSATPKQKIHNF
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  SHLVSKOELLELFELMTGRXKDGQLTGVLGVYFWVQKSTINTIMONKVSVSATPCH
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  TWFPQVLVERGLCLDCPLWWSFVSTAELWTCGLIPIDQWHDVFPVSLVQWIPR
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  HVLKATYGINIITPREDEPHRPTSEELLTATGTHRGYNTANGVQFPASANTILADTV
SEG  .....
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  SKLLYCHPPFGROPPTTFQWQWLLLENHWSDELTKQLGRNKAQIEMIVQKTFHQE
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  MVAALTKGVQVWGYTEPGSOVVTASTASSEHGAGKFWXKGNRHKESRALTENLW
SEG  .....
PRD  hhhhhhhhecccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Posite for OXfphes3_Be24.3

PS00001	264->268	ASN_GLYCOSTYLATION	PDCC00001
PS00001	359->363	ASN_GLYCOSTYLATION	PDCC00001
PS00004	410->414	CAMP_PHOSPHO_SITE	PDCC00004
PS00005	21->24	PKC_PHOSPHO_SITE	PDCC00005
PS00005	28->29	PKC_PHOSPHO_SITE	PDCC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDCC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDCC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDCC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDCC00005
PS00005	493->496	PKC_PHOSPHO_SITE	PDCC00005
PS00005	531->534	PKC_PHOSPHO_SITE	PDCC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDCC00005
PS00005	649->652	PKC_PHOSPHO_SITE	PDCC00005
PS00006	52->56	CK2_PHOSPHO_SITE	PDCC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDCC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDCC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDCC00006
PS00006	155->159	CK2_PHOSPHO_SITE	PDCC00006
PS00006	259->266	CK2_PHOSPHO_SITE	PDCC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDCC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDCC00006

WQ 01/12659

PCT/IBRU01496

PS00006	281->285	CK2_PHOSPHO_SITE	PDCC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDCC00006
PS00006	299->303	CK2_PHOSPHO_SITE	PDCC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDCC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDCC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDCC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDCC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDCC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDCC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDCC00006
PS00006	505->509	CK2_PHOSPHO_SITE	PDCC00006
PS00007	480->488	TYR_PHOSPHO_SITE	PDCC00007
PS00007	190->198	TYR_PHOSPHO_SITE	PDCC00007
PS00008	9->15	MYRISTYL	PDCC00008
PS00008	432->438	MYRISTYL	PDCC00008
PS00008	620->626	MYRISTYL	PDCC00008
PS00009	1->5	ANIGATION	PDCC00009
PS00009	378->382	ANIGATION	PDCC00009
PS00017	393->401	ATP_GTP_A	PDCC00017

(No Pfam data available for DRF2phtes3_Ba24.3)

DRFphes3_8g11

group: testes derived

DRFphes3_8g11 encodes a novel proline-rich 919 amino acid protein without similarity to known proteins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop).
No informative BLAST results; No predictive prosite, pfam or-SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

1 EST hit (from testis library)

Sequenced by MedGenomix

Locus: unknown

Insert length: 3100 bp
Poly A stretch at pos. 3056, polyadenylation signal at pos. 3041

```
1 AGAGTCTTCC CTCAGCATAT TTTACGATAG AGAGATCTTT GTTCCAATGG
51 AAGAAAGTGA GGAATCAGAG AGTATTCTCC AGACAGAGAT TTCTGAGTCC
101 CACACTCTCC TCAGCTCAAA TTATCTTTCC CAGGCTCAAG CTGACTCTTC
151 AGACAGCTTC CAGTTGCTAG AGATCTCGCA GCTAANAATA GAGCAAAAAC
201 TCTTAAGGAG TCAATATACC CCGGATGTCC CTCACCTCTT AGTTCAAGGT
251 CTGCTCTTAA AATACCTTAT CTGCTACAG TTGTGCCAT CTTCAGAGAT
301 TAATTCCTAT CATAAATTAC AGACCACTTC GGAGCCTTAT CTCTTTATCT
351 ATCLACAGCT CGACTCTGTA GGCATCTCTG AAGGCTATGG TGAAGTCTGG
401 TTGATCTTTC GCTTTAGCT GAGATTTGGG AAGAGTCTCC AATCTCAAA
451 GTATCTGTGA AGAGATAGAC CCGTCATAGC GAGAGGCCCT ATATCAACAT
501 CACAGAGGAA AGCTAAATAT TATAGTCAAG CTTCAGAGC TCCATCTTCC
551 ACAATGATT TGCATCTGG GCTTCTCCAG TCCCTGCTC CTGTACAAGT
601 CTACATCAGG GAGGACAGAG GAGCAGAGCC TACCTTAGTA GAAGAAGAAA
651 AACTTAGAGC ACTTGGGCTC TATCAATTCA CTCAGCTTCA CAACCTTCCA
701 GAGAGTCACT CTGAAGGCTC TCAGATGAAA AAGCTGGCTA AAGTGAAGAC
751 CAAGAAGAGC TCTGATTCCA AATATGCAAT GAAGAGATCT ACCAGGAGAC
801 TTGAAGAACA CAGGAAGTTC TACACAAACA GTAGAACAC AATAGAGAGT
851 CTTCTAGGGG AATTAGAGC CCATTTAAGA AGGAAGAGGA TTGAGGCAAC
901 CTGACAGAGT AGTCTCTTTT TAAAGAACA ACTTAAGAAA CTTTCCACAC
951 CGAAGTTCAT GCAACTCTTT TTTAGAGGCC TAAAGCGGGC ATTCCAAACA
1001 GCACACAGAG TTATAGCTTC TTGTGGGGGG AAGCTTTGAG ACAGGACAGC
1051 GGCACAGAGT TTGTGGGCTA GCAAAAACCTA TTATCAAAA CAAATGCCCA
1101 GGAATCTATG CTTACCAAGC AGTATCAAAA GAGACAGAG CTACAGCTGAC
1151 AAGCTACGCG CAGCAGGCTC AACCAATTAG CAGGAGGACA TATTTGTGGG
1201 AGAAGGGTCC CAGTGCAGAT CAGCTCAACA GCAAGAGAAA GCTTACTCTT
1251 TCCAACTCCG ACTCTTTGCA TTGCCCCAGC CCACAGATTC CCAAGTGTGT
1301 ATGCTTTTCC AACTTCTCTC AGTGGGGGAG CCTCTAGAAA CTCTCAAAA
1351 GGCAGAGTGT AGCAGATCAA AGAAAACCTT CTATAGAAAT GAAGCTTCCA
1401 GGCAGAGATC TAGAGACTTC TCACACAGAG GAACAGAGAT TCAAGGTCCA
1451 GGAAGATCCG TACTGTGTTT CCGTGTGAAG AGACCTGGGC ACCGACATCT
1501 TAAAGACAAA CTCACACACA AGGAGCATAA CCACTCCAGC TTCTATAGGG
1551 AGCAACCTCC AGGCTCTCTT TCTGAGAGAA CCGCTCATAA CCTCTCTGG
1601 AGAACCATCC GCACTCTCTC TGAGAGAGAG CAGCCAGATT CTTTGGAGAG
1651 ABAACATCAC AGTCTCTCTC AGAGAGAGCA CTGCATCTCC TTAGAGAAA
1701 AGATCTCCAG TCTCTCTGAG AGAAGCTGGC GCAGTCCGTC TGAGAGAAAT
1751 CACTGCACTC CCCCCGAGAG GAGCTGTGAC AGTCTCTCTG AAGGGGGGCT
1801 TCACAGTCTC TCACAGAGGA GGCATGCGGG TCCTCTCTAG AGACATATC
1851 ACAGTCCCTC AGAGAGAGC CATGCACTC CTTGAGAGAG AAGCCATCTC
1901 AGTCTCTCTC AGAGAGACA TGCAGTCTCC TCCAGAGAGA GCTATCTGGG
1951 TCTCTCAGAG AGAGGCATTT GCACTCTCTC TGAGAGAGGA CATCTGCTTC
2001 CTTCTCAGAG GAGCCTCTGT GCTCTCTCTG AGAGAGAGCA TCACATCTCC
2051 TTTAGAGGAA GCAATGCTCC TCCGCTGGGG AGAGCCTAT GCACTCTCTC
2101 AGAGAGAGAG CATCACAGTC CTTCTGAGAG AAGCCATCAC AGTCCCTCTG
2151 AGAGAGAGCA TCACATCTCC TCTGAGAGAA GCAATCTGAG TCCCTCTGAG
2201 AGAGCATTT GCACTCTCTC TGAGAGAGCA CATGCACTC CTTCTGAGAG
2251 AAGACATCAC ACTCTCTGAG AGAAGAGCCA TCACATCTCC TCTGAGAGAA
2301 GCAATCACAG TCTCTCTGAG AGAAGAGCTC ACATCTCTCT GAGAGAGAGC
2351 CTTCTCATCT TCTCTGAGAG GAGCATCTCC ACTCTCTCTG AGAGAGATCT
2401 TCACAGTCTC TTTGAGAGGA GCAATCTGAG GATTCTCTAG AGAGTCAACA
2451 TTCTCTCAGA GAGAGAGCAG CTGAGTCTCT TGGAGAGAG CCGTCTGAGT
2501 GCTCTCTGAG GAGAGAGACA CAGTCTCTCT GCGAAGACTC CTACAGCTCC
2551 CTTCTGAGAG AGCATCTGCA GTCTCTCTGG GATGAGGCAA GCGAGGAGCT
2601 CTGAGAGAG GATCTGAGAT TCTCTGAGA GAGCTGCTCA CATCTCTCTT
```


WO 01/12659

PCT/IB00/01496

```
2451 GAGATGAGGC CAGGAGGGCC CTCGGGAGG AACCATGCA GTCCCTCTGA
2701 GAGAGAGCGA CCAAGTCCCC CTAGAGAGG ACTCAAGAC AGTTCTCTGT
2751 GAGAGAGGCC CAGCCTTCT TTCTCTAGG ATTTCAAGAA TCACACACT
2801 CTCCTCGGGA CACACATATA AATCCCAA GCAGGCGAG TGTGAGGCC
2851 TGAGGTACT CAGTGAAGGC AGTCCGCC CATTATCTA TTCTCTTAG
2901 TCTTCATCGT GGTGCCCTTT CAGGCTTCT TTCTCTCTCA GCACCTGCT
2951 CCAATTCCTG CCGCCCGAGC GTGCAAGGC TTCCATTCT CTCTACCGGG
3001 GGGAGAGGGC CTAGCAATGC CTCTCTTAAT TTCTTAGAT GAATAAAGG
3051 GCAGTTAATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 2863 bp; peptide length: 939
Category: similarity to unknown protein
Classification: unclassified
Prosite motif: ATP_GTP_A (1824-832)

```
1 HEESDGGSD SVPRISEGH ELAPHYLGA HTDSEQOL LEDIOLRIA
51 ELASDIPPO VPPPLAGLV LAFPLCQG KCSGLMCHK LQTSQPTLL
101 IYPOHLVRY PEHGCDFVLM LAFPLAIGR SQIRYNERD RPVIKSPIS
151 PGRMAKITT QHSKSTFTI DUGSGRSGR HPQWVIRHC ORSMOLVEX
201 TETRAAPGIE FTQVHMLPS DSESTORER AKVNTKTSO SKYHWRITE
251 RLKXKRPFT MHTTIESPS RELAAHLRK RIGATOTSTA SLAKPKPXS
301 OPTTOLLITO SLRATCTAR WILASVGRP VICTTFOMLA ASSTYFQCP
351 ADYCLP331 KDKRBAKIL TTAGSTIQE DILMGTVOC NSAQPPRAY
401 STPHMLAPL APTDGGGTA TGTASVQGL RVVQDSESR SEKAFARET
451 SQSEKMLST PSTRVQAGR ILGSIYERT WHNLEKRLT HXENHPSFT
501 KERTFNGPSE HTKHPMSHM HSPSPSPQR SLEPRHHSP SQKNCSPFR
551 ENKSPSPERS HSPSPSPHC IFFERKIEL HENGLHPSG RHMCPSPDM
601 HNSPSPERSH SPERSHSPR DEKRNHSPQ RSHNGSPERS HCSFSPERNH
651 SPSPSPSPR SPERSHSPR HNSPSPERS HNSPSPERSH SPSPSPSP
701 SPERSHSPR HNSPSPERS HNSPSPERS SPERSHSPR SPERSHSP
751 SPERSHSPR HNSPSPERS SPERSHSPR SPERSHSPR SPERSHSP
801 SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSP
851 SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSP
901 SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSPR SPERSHSP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DEF08253_0g11, frame 2

TRNDL:AF061185.1 gene: "car90"; product: "cyst germination specific
acidic repeat protein precursor"; Phytophthora infestans cyst
germination specific acidic repeat protein precursor (car90) gene,
complete cds., M = 1, Score = 457, E = 2.3e-39

TRNDL:AC004361.38 gene: "F1492.41"; product: "putative proline-rich
protein"; Arabidopsis thaliana chromosome 11 BAC F1492 genomic
sequence, complete sequence., M = 1, Score = 340, E = 4.2e-27

TRNDL:AF062653.1 product: "plenty-of-proline-101"; Mus musculus
plenty-of-proline-101 mRNA, complete cds., M = 1, Score = 313, E =
3.6e-24

PIR:P00099 son) protein - human (fragment), M = 1, Score = 292, E =
1.2e-22

>TRNDL:AF061185.1 gene: "car90"; product: "cyst germination specific
acidic repeat protein precursor"; Phytophthora infestans cyst germination
specific acidic repeat protein precursor (car90) gene, complete cds.

Length = 1,489

HSPa:

Score = 457 (68.6 bits), Expect = 2.3e-39, P = 2.3e-39
Identities = 91/444 (20%), Positives = 239/444 (53%)

Query: 475 SPVKTWIRHLKDLTKKHNPSTY-RETPRGPSERTRHNPFWNHNSPSEASQSSL 533
+P + T + + + T + + + E T P + E T + P + +P + + +S
Sbjct: 584 APTEETNIAPIEET-TYAPTEETTYAPAGEETVPEPEETTYAPTEETTYAPTEETIAT 642
Query: 534 ERNHSPOQRHNSPSEKHSPPSEKHSPOQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHS 593
E + +P + + +P + + P + E + +P + + +P + E + + +E + +P + + +
Sbjct: 643 ECTTYAPTEETTYAPAEETPIETPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 702
Query: 594 RQSPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 653
P + + + P + E + +P + E + +P + E + +P + + GP + E + +P + E + +P +
Sbjct: 703 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 762
Query: 654 QRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 713
+ + P + E + +P + + +P + + +P + E + +P + E + +P + E + +P + E +
Sbjct: 763 ECTTYAPTEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 822
Query: 714 CSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 773
+P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 823 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 882
Query: 774 ERNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 832
E + +P + + +E + + +E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 883 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 942
Query: 833 HSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 892
+P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 943 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1002
Query: 893 KEGKTSFPGERPSSHLSDPKHOTT 918
+E + Y + P + E + + + + + + T
Sbjct: 1003 EE-TTYA-PTAEETTYAPAEETTYPEET 1026

Score = 445 (64.8 bits), Expect = 4.5e-38, P = 4.5e-38
Identities = 83/394 (21%), Positives = 212/394 (53%)

Query: 502 EATPRGSPERTRHNPFWNHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 561
E T P + E T + P + +P + + +E + +P + + +P + + +P + E + +P + E +
Sbjct: 763 ECTTYAPTEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 822
Query: 562 RQSPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 621
+P + + P + E + + + +E + +P + + +P + + +P + E + +P + E + +P +
Sbjct: 823 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 882
Query: 622 ERNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 681
E + +P + + +P + E + +E + +P + + +P + E + +P + E + +P + E + +P +
Sbjct: 883 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 942
Query: 682 RQSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 741
+P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 943 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1002
Query: 742 ERNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 800
E + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 1003 ECTTYAPTEETTYAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1062
Query: 801 HSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 860
+P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 1063 YAPAEETTYPEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1122
Query: 861 EPTNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 894
E T + +P + E + +P + E + +P + E + +P + E + +P + E + +P + E + +P +
Sbjct: 1123 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1186

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
Identities = 84/421 (20%), Positives = 223/421 (52%)

Query: 475 SPVKTWIRHLKDLTKKHNPSTY-RETPRGPSERTRHNPFWNHNSPSEASQSSL 533
+P + T + + + T + + + E T P + E T + P + +P + + +S
Sbjct: 848 APTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 906
Query: 534 ERNHSPOQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 593
E + +P + + +P + + P + E + +P + + +P + E + +P + E + +P + E + +P +
Sbjct: 907 ECTTYAPTEETTYAPAEETPIETPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 966

969

970

971

972

Sbjct: 997 TWAPIEETTYAPTEETTYAPAEETPIETETTYAPTEETTYAPTEETTYASTETTYA 1054
Query: 652 PSQRSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 711
P++ + P+E + P++ + P+ + P+G + ++P+E + ++P+E ++P+E
Sbjct: 1057 PTEETTYAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 1116
Query: 712 SHCSPEKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 771
+ P+E + ++P+E ++P+E ++P+E++ + P+E + ++P+E ++P+E + ++
Sbjct: 1117 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1176
Query: 772 LLESHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 830
E + P++ E + + E ++P+E++ +P+E + P+E ++ + +
Sbjct: 1177 PTEETTYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYPIETETTYAPTE 1234
Query: 831 TCHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSH 890
T + P+E + ++P+E ++P+E + E T ++P+ + P+ ++P+E + +
Sbjct: 1237 TTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1294
Query: 891 PLEE 894
P +E
Sbjct: 1297 PTEE 1300
Score = 403 (60.5 bits), Expect = 1.6e-33, P = 1.6e-33
Identities = 84/104 (21%), Positives = 213/334 (54%)
Query: 501 RERTPGSPERTKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 560
RE T + P +P+E+ +E + ++ +P++ ++P+E
Sbjct: 319 RERTYAPAEETTYAPAEETTYAPTEETTY--DVEETTYVTEETTY--APTEETTYAPTE 375
Query: 561 WSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 620
+ ++ C E + ++ +E ++P+ + P++ ++P+E + P+E + ++P
Sbjct: 376 HYAHENP-COT-DVTHAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 433
Query: 621 SPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 680
+E +P++ + P+E++ +P+E +P++ + P+E ++P+E + P+ +
Sbjct: 434 YEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 493
Query: 681 HNSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 740
+ +E + ++P+E + ++P+E + P+E + ++P+E + ++P+E +P+E +P
Sbjct: 494 TYASTETTYAPTEETTYAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTEETTYAP 553
Query: 741 SPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 799
+E++ ++P+E + P+E ++P E + ++ E + +P E ++ E + + E
Sbjct: 554 TEETTYAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEET 613
Query: 800 SHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 859
+ P+E++ +P E + ++P+E ++P+ ++P+E + +P+ +E + +
Sbjct: 614 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYAP 673
Query: 860 CERTKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 894
E T ++P+E P+ ++P+E + ++P+E
Sbjct: 674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 708
Score = 398 (59.7 bits), Expect = 5.5e-33, P = 5.5e-33
Identities = 84/102 (20%), Positives = 209/402 (51%)
Query: 475 SPVKTWHKILKILKILKILKILKILKILKILKILKILKILKILKILKILKILKILKILKIL 533
+P + T + ++ T+ + E TP P+E T + P+ +P+E + +S
Sbjct: 992 APTEETTYAPAEETTYAPTEETTYAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTE 1050
Query: 534 ERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSH 593
E ++P++ + P+ + P+E + P++ +P E + ++ +E ++P++ +
Sbjct: 1051 ZETTYAPTEETTYAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
Query: 594 RQSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 653
P++ + P+E + ++P+E + ++P+E +P + + GP+E + ++P+E +P
Sbjct: 1111 YAPAEETTYPIETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1170
Query: 654 ORSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSH 713
+ + P+E + P+ + P+ + ++P+E + ++P+E +P+E +P+E +
Sbjct: 1171 EETTYAPTEETTYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYPIET 1230
Query: 714 CSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSP 773
++P+E + P+E ++P+E ++P+E++ +P+E + +P E + ++
Sbjct: 1231 YAPTEETTYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1290
Query: 774 ERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSH 833
E + ++P+E + E + P+ ++ +P E + ++P+E ++ +T +
Sbjct: 1291 EATTYAPTEETTYAPTE-----ETTYPIETETTYAPTEETTYAPTEETTYAPTEET 1343
Query: 834 SPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 896

974

Report for DEFPptes3_8q11.2

975

PCT/IB00/01496

```

Promote for DKFzphes3_gll.2
P500017 639->647 ATP_GTP_A P0C00017
(No Prim data available for DKFzphes3_gll.2)

```

WO 01/12659

PCT/IB00/01496

DEPhtes3_Bp5

group: cDNA derived

DEPhtes3_Bp5 encodes a novel 166 amino acid protein nearly identical to human KIAA087 protein.

The novel protein is a new splice variant of KIAA087.
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

KIAA087, alternative spliced

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
No poly A stretch found, no polyadenylation signal found

```

1  CCAGACATCGG  CGGTGTCTCC  AGCAGCTGCC  GGGGGTGGCG  CGAGCTGTGC
51  CAGAGCAGCG  GGAAGGTGTG  GAAGGAGGAG  TTCCGGGTGA  GGTGACCTTC
101  CTTATATGAA  CACTACAGCC  CAGCGGACTA  CGTAAATGCG  TTGGAGAGAT
151  ATAACTGTGG  GCAAAAGAGT  GGGTTAGAGG  GGGGGAAGAT  TGTAGCTCTG
201  TTCTCAAGAA  GGTCTCTTTC  AGAGCAGCTT  CCTTGTAATG  GTCTTAGTGA
251  CATTCGAGAG  CTTCGAGGAG  CAGACATTTT  TTTCGAGGAT  GAATCTGGTT
301  GTATCTCTAA  TATGGAAGGA  AGAAAGCTTT  TGACCTGGAA  ATACTACCGA
351  AAAAAAATTC  TTACTTCTCT  GGGGCAAGAG  AGATCTCTAA  ATATCTCTAA
401  GGGCTTCTCT  CAGGAGCAG  ATGACTATGA  GTCTATCTTT  GAAGGTCTGT
451  TATATATTTA  CAGTACTTGC  AATCTCTCTT  CCGACATCAG  GGTCAAGAGC
501  ATTCAGGGCC  AAATTGACAG  CATCTGGGAG  CTCTCTTGGA  AAACCTCTCG
551  GGGCATTAAG  AGTGGCAGCC  CAGGCTTGGC  CTCAAGGCGA  GGTGAATCAT
601  GCATATATAT  GGAATAGAGA  CTCCAGAGCG  AGGTGCTGGA  TGCCATGAAC
651  TATGTCTCTT  AGCAGCAACT  GAAGTTCAG  GGAATCGAA  TGCATTACTA
701  TAAATGCCCTC  AACTATATTA  TGCATCAGTT  TTGATCTGCG  AGAGCAGGAA
751  TCCCAATGAG  CATCTCTCTG  CTCTATTGGA  CAATCTCTCG  GCAATCTGGA
801  GTCCCACTGG  AGCTGTCTAA  CTCTCCAGAT  CACTCTCTAT  TAAGGTGATG
851  CCAGAGGCGA  GAAGGGGCGA  CCGTGGACAT  CTCTGACTAC  ATCTATATAG
901  ATGCTTTTGG  GAAGGCGAAG  CAGCTCAGAG  TGAAGAATG  CGAGTACTTG
951  ATGGGCGAGC  AGCTGACTTC  AGCACTGTAT  GGGGTGCTCA  ATCTCAAGAA
1001  GGTCTTACAG  AGATCTGTCG  GAACTCTGTT  AAGCTCTGGG  AAGCGGAAG
1051  GCATCGACCA  GTCAATCCAG  CTCTCGAGAG  ACTCGCTGGA  TCTCTATCTG
1101  GCATGTAGC  GGGACAGGCT  GGAATCTCT  CTCTCGAGAG  CGAGCTTTTA
1151  CTCTCAGCTG  GGAATCTGCG  CAGAGAGGTC  TTCTCTCTTT  GTTTTGAGGG
1201  TGTCTGACAT  CTCTCAGCAC  ATCCAAATCC  TAGACCGGGG  GCAGCAGGGG
1251  GCGTGGGCT  ACCTGTCTGA  GCACACTCTT  GAGACATCTG  AGCGAAGAAA
1301  GGAGGAGGTG  GGGCTAGAGG  TGAAGCTGCG  CTCTGATGAG  AAGCAGCAGG
1351  ATGTGTGCTA  CTGATCTGGG  CTCAATTATG  AGCATAGAGG  GTATGCTGAT
1401  AACTGTGTGA  TGTAGCGCTC  GGAACCCAGC  TGCATATGCG  GACACAGCTG
1451  GATCCGGAGC  ATGAGCTCTC  ACAGCTCTCG  GCAGGGGAGC  CACAGAGCTT
1501  TCTATAGCTT  CTTGTGGGAG  GAGGAGCTCT  GTGGATAGCG  AGCCGAGGAA
1551  AACTTGGAA  ATAGCTGTGA  GCTTCAGAGA  ATCTACACAC  CTGACGTGGG
1601  AGCTTATCT  TGAAGCTTTA  CTGGCACTCT  CTACATCTGA  AAGCGAGAGC
1651  TGGAGATCTG  GTATCCAGAA  GATCTGGAGT  TTGTCTATGA  AACGCTCGAG
1701  AATATTATCA  GTCCAGAGAA  AGAGACATTA  GATGATTAAG  GTGTGAGAG
1751  GACATTGAC  CTTTCTGCTT  GCTGCTATCT  TCCAGAGAAA  CGGACTCTCG
1801  GAAGAGAGAG  TCTCAGCGGA  GCGCTCGGGA  CTTCTGTCAG  CAGGAGAGGC
1851  ACTCCAGCAG  TACTCTGCTT  TCTCTCTAC  TAACTTTTAA  TACCTCTGCG
1901  TCTTCCCGAG  CTCCAGAGAC  AATCTGTCT  TCCGCTTACA  CTAGTGAATT
1951  AATCTGAAAG  GCATCTGTCT  AGTGGATGCG  CTGTATGCTT  TGTCTGTGG
2001  TGACAGTTTG  TGACATCTTG  TCTTCATGAG  GTCTCAGAGT  CGAGCTCTCT
2051  GTAATCATTC  TTGTATTTCA  CTCCATCTGC  CTCTCTCTCT  GCAATTTCTT
2101  CAGACATCTT  CTTCTGCTGG  ACAGATGGGG  TTATGATCTT  GCAATATTTT
2151  CTTCTGATTT  TCTCTGTGGA  AGTGTCTTGG  TCCGAGTGA  GCACTGTGTT
2201  TCTTTTAC  CTGATTTAG  TTGATATTT  AGAGTAAAG  TTGTGTGCTA
2251  TCTTGGCAGC  ATCTTACAGA  TGGAGACATT  AAGAGCTTAA  TGTATTATAG
2301  AATCATTTGA  ATTTATTTTT  TCTTAATATG  TGAAGCAGAG  ATTTCAAGTG
2351  TTTATCTT  TTTTCTTT  AATTTAATG  GGAATATAG  AGCTTTTCT
2401  CTTCATATT  CTTCTCTTGA  GTTTATGAG  ATCTCTATTA  ATCATGATT
2451  TCTATTTTA  TTATATAAA  TCTTTTGA  AATGCAAT  AGTGACTTT
2501  GTCAATGAT  TTTTCTATC  TCACTAGAA  TTCTCTCAT  TTAATGACT
2551  ACTTTTATT  TTTATTTAA  AAAATCTACT  TCAATATCAT  GAGTACTCT
2601  TACATCAGTG  ATGGTCTCT  TTTGTAGTGA  GACATACAAA  TCTATCTTA

```

WO 01/12659

PCT/IB00/01496

2651 ATGTTTGGCTC TTAGAAGTCA TACTCATG TCCTCAAGA CCAAAAAATG
2701 AGGTTTGGCTC TTCTTAATCA GGAAGAAAAA AATTAAATGA CTTAAAAA
2751 AAAAAAAAAA GG

BLAST Results

No BLAST result

Headline entries

No Headline entry

Peptide information for frame 3

ORF from 105 bp to 1736 bp; peptide length: 516
Category: unknown protein
Classification: unclassified

1 MNKYSPTDVI MWLEEVVRO KAGLEAKIV ASFRKFFSE HVPKGFSDI
51 EDLGVLEIFF EDLVCILMH KGSALTWET YAKSILVLA QDILMHLA
101 FLOQPDYES YLEGAVTIQ YCNPLSDISL KQIQADSI VELVCKTLNG
151 INSRHSLAF KAGESHNE ELQSOVLDA HWVLYDOL FQHMWTH
201 ALMLHNMCI IRTGIPISN ELLTITAND LQVLEPWF PSFLLMNCQ
251 GAGGATLDIF DYIIDAPEK GQLTWKEE YLIGHVTAA LYGVWVKEV
301 LQNVGHLLS LGRREGIDQ YQLLAEIDL LAMYPDQV LLLIGARLF
351 HLGIMPEKSF CLVLKVLDEL QNTOTLQDQ NGAVGYLQV TLKHIERKE
401 EVGVVLEAS QSKNMCHIS LLLIKAKAT CMCWITQD PTCHKHEDI
451 KNDHNSLPM QNDQPTFVL PEDLEPVET VQHLTSKKE WIDE
501 YPSEPTGTH IPRAELEIRY PEDLEPVET VQHLTSKKE WIDE

BLAST hits

No BLAST hits available

Alert BLAST hits for DEFspres_6q5, frame 3

THREMLMCH:AB020682.1 gene: "XIAA0875"; product: "XIAA0875 protein";
Homo sapiens mRNA for XIAA0875 protein, partial cds.; M = 1. Score =
2832, E = 5.5e-295

>THREMLMCH:AB020682.1 gene: "XIAA0875"; product: "XIAA0875 protein"; Homo
sapiens mRNA for XIAA0875 protein, partial cds.
Length = 621

HSPs:

Score = 2832 (424.9 bits), Expect = 5.5e-295, E = 5.5e-295
Identities = 537/544 (98%), Positives = 537/544 (98%)

Query: 1 MNKYSPTDVI MWLEEVVRO KAGLEAKIV ASFRKFFSE HVPKGFSDI EDLGVLEIFF 60
Subject: 85 MNKYSPTDVI MWLEEVVRO KAGLEAKIV ASFRKFFSE HVPKGFSDI EDLGVLEIFF 144
Query: 61 EDLVCILMH KGSALTWET YAKSILVLA QDILMHLA FLOQPDYES YLEGAVTIQ 120
Subject: 145 EDLVCILMH KGSALTWET YAKSILVLA QDILMHLA FLOQPDYES YLEGAVTIQ 204
Query: 121 YCNPLSDISL KQIQADSI VELVCKTLNG INSRHSLAF KAGESHNE ELQSOVLDA 180
Subject: 205 YCNPLSDISL KQIQADSI VELVCKTLNG INSRHSLAF KAGESHNE ELQSOVLDA 264
Query: 181 HWVLYDOL FQHMWTH ALMLHNMCI IRTGIPISN ELLTITAND LQVLEPWF 240
Subject: 265 HWVLYDOL FQHMWTH ALMLHNMCI IRTGIPISN ELLTITAND LQVLEPWF 324
Query: 241 PSFLLMNCQ GAGGATLDIF DYIIDAPEK GQLTWKEE YLIGHVTAA LYGVWVKEV 300
Subject: 325 PSFLLMNCQ GAGGATLDIF DYIIDAPEK GQLTWKEE YLIGHVTAA LYGVWVKEV 384
Query: 301 LQNVGHLLS LGRREGIDQ YQLLAEIDL LAMYPDQV LLLIGARLF HLGIMPEKSF 360
Subject: 385 LQNVGHLLS LGRREGIDQ YQLLAEIDL LAMYPDQV LLLIGARLF HLGIMPEKSF 444

WO 01/12659

PCT/IB00/01496

Subject: 385 LQKRVGVLSSLRKGGIDQYLLADSLDLYLAMYPOQVQLLLQARLYPHGIWPEK-- 442
 Query: 361 CLVLAVLDLQHIQTLPQGGAGVLYVQTLNHTPKKEVGVEVLASDEKHNDVCS 420
 Subject: 443 -----VLDLQHIQTLPQGGAGVLYVQTLNHTPKKEVGVEVLASDEKHNDVCS 497
 Query: 421 IGLIMHKRYTMCVYICNDPTCMGHEWIRRMQVSLPHGHQPFYVVLVEGSCRYAA 480
 Subject: 498 IGLIMHKRYTMCVYICNDPTCMGHEWIRRMQVSLPHGHQPFYVVLVEGSCRYAA 557
 Query: 481 QKLEYHVEPQEISHFVGATYFSEFTGTHYIYMAELEIRYFEDELPVYETVQHTSAKKE 540
 Subject: 558 QKLEYHVEPQEISHFVGATYFSEFTGTHYIYMAELEIRYFEDELPVYETVQHTSAKKE 617
 Query: 541 HIDE 544
 Subject: 618 HIDE 621

Pedant information for DRFiphes3_Rq5, frame 3

Report for DRFiphes3_Rq5.3

[LENGTH] 544
 [MW] 63307.22
 [pI] 5.82
 [CDMW] YIPDL:AB020482.1 gene: "KIA0875"; product: "KIA0875 protein"; Homo sapiens
 KMA for KIA0875 protein, partial cds. 0.0
 [FW] Alpha Beta
 [FW] LOW COMPLEXITY 1.84 s

SEQ HNRHSTPTQVWVLEEVQKAGLEAKIVASFKRFFSEIVPCNGFSDIENLEGKLIFF
 SEG
 PRD ccc
 SEQ EDELVCILMRECKALTWYFAKILYLAQQLIMWLKAPLQKQDYTESYLEGAVIIDQ
 SEG
 PRD ccc
 SEQ YCPFLADISLQDQIDSVLEVCHTLAGTMRNPPLAFYAGESSHIMELQSQVLDA
 SEG
 PRD ccc
 SEQ HNVVLYQLKFGKMGDTYMAKLYNHQVLIARTGIPISMSLLYLTIAQLQVPLEPVWF
 SEG
 PRD hhhhhcc
 SEQ PSHFLRMQCGAGATLDIPDYIYDAFGKGLTVKCEYLIQHYTAALQVWVWLVV
 SEG
 PRD ccc
 SEQ LQKRVGVLSSLRKGGIDQYLLADSLDLYLAMYPOQVQLLLQARLYPHGIWPEKSF
 SEG
 PRD hhhhhcc
 SEQ CLVLAVLDLQHIQTLPQGGAGVLYVQTLNHTPKKEVGVEVLASDEKHNDVCS
 SEG
 PRD kdnnd
 SEQ IGLIMHKRYTMCVYICNDPTCMGHEWIRRMQVSLPHGHQPFYVVLVEGSCRYAA
 SEG
 PRD ccc
 SEQ QKLEYHVEPQEISHFVGATYFSEFTGTHYIYMAELEIRYFEDELPVYETVQHTSAKKE
 SEG
 PRD hhhhhhhcc
 SEQ HIDE
 SEG
 PRD GECG

(No Prosite data available for DRFiphes3_Rq5.3)

(No Pfam data available for DRFiphes3_Rq5.3)

DEFspices3_0ml0

group: nucleic acid management

DEFspices3_0ml0 encodes a novel 221 amino acid protein with strong similarity to polyadenylate-binding proteins.

The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.

The new protein can find application in modulation of mRNA translation and processing/stability.

strong similarity to polyadenylate-binding protein

frame shift at Rp 707-710

Sequenced by MediGenomix

Locus: unknown

Insert length: 2107 bp

Poly A stretch at pos. 2052, polyadenylation signal at pos. 2033

```
1 CGGAAAGGTC GCGGCTTGTG TGCCTGGGG CAGCCTGCG GAGAAAGAAC
31 CCGAGAGACC CCGAGTACGC AAGGAGCTGG CTCCTGCTGG GGGAGCTCCA
101 CCGCAGCTG ACTGAGGCGA TGCCTACGA GAAGTTCAAG CCGGAGGGGC
151 CATCTCTTC CATCGGATC TCGAGGACT TGAATACAG GGGCTCTCTC
201 AATGAGGCT ATCTGAGCTT CCGAGTACGC AAGGAGCTGG AGCATCTCTG
251 GAGCAGCTG AATTGTGATG TTATAAGGG CAGGACAGT GCGATCATGT
301 GGTCTCAAGG TGATGCTCA CTCGAGAAA GYGAGTGGG CACATATTC
351 GTTAAAGTC TGGATAGTC CATTAAAT AAAGCACTGT ATGATACAGT
401 TCTGCTTTT GGTAACTCC TTCTGTATA GGTGCTTGT GATGAAATG
451 GTTCAAGGG TTATGATTT GACACTTTG AAGACAGCA AGCAGTTAA
501 AGAGCTATTA AAAAATGAA CCGAATGCT CTAAATGCT CCAAGATATT
551 TGTGACAAA TTAAATCTT CTAAAGAGC AGAAGTCAA GTTGAGCTA
601 GGGCAAGCA GTTCCCAAT GTTACATCA AGAATTTGG AGAGCAGTS
651 GATGATGAG GCTTTAGCA TCTTTTGGG AAGTTGGCG CCGCTTTAG
701 TGTGATTA TGCAGCTCA AATCTGAAA TCGAAGGAT TTGATTTGT
751 AGCTTTTGA AGCATGAGG ATGCACAGA AGCTTAGAT GAGATGATG
801 GAAGGAGCT CAGTGAANA CAATTTTGG TTGCTGAGC TCGAAGAAA
851 GTGAGAGCG AGACGAACT TANGGACAA TTGACAGA TGAGCAGAA
901 TAZGATCAG AGATACAGG TTCTTAATC TTATGAGA AATCTGATG
951 ATGTATTGA TGATGAGCT CTGGGAAAG CATTTTCTC AATTTGATCA
1001 ATCACTAGT CAAAGTTAT GATGAGGCT GTTGGAGCA AAGGCTTTGG
1051 TTTTGTATG TTTCTCTCC CAGAGAGCG CACTAAGCA GTTACAGAA
1101 TGAGCGTAG AATTGTGGC AGAAGGCAAT TGTATGAG TTTAGCTCAG
1151 GCGAAGAA AGGCTGAGG TTAGCTACT AAGGTATA TGGAGGAT
1201 GCGAGTGA CCGGCTGTC CCAAGCAGG AGCACTCTT TCAGTTACT
1251 TATAGAGC TGTCACAGC ACTGAGACC ATGCTGCATA CTATCTCTT
1301 AGGCAATG CTCGACTAG AGCAATGCT CCGTGACTS CTCAGGTTG
1351 CAGACCTAT CATTCCAAA ATAGGCCAG TCGTATCCG CCGAGTCTC
1401 TAGATGAC AATTAGTACT ATCAGAGAG CTTCTGACA GTTTGAGAA
1451 GTCATGTCA CCGAGCTGT TGTAAACA TCAAGACAG CAGTGGTCC
1501 AGTCTCTGA GCTGCTGCT GTGCTGAGC TACGCTCTT GTGGGAGGG
1551 TCCAGGTA TAAATATCT GCGGAGTTC GCATCTCTA CCAAGATCT
1601 ATGACAGC CACAGTTAC AATGACAGG CTGCTGTTT ATGTACAGG
1651 TCGAGAACT TGACCTGCT CAGGTTGGC ATGCTGCTT CTTAAAGG
1701 AAAGCAAT GTTAGTGAA CCGCTCTTC TCTTATTA AGCATGAC
1751 CTAATCTTT GTGGAAAT CACTGAGAT TGTGAGAA TGTATATTC
1801 AGAATCTTT TATATCTTC AGCTTCAGA CTCACTGCT TTAAGTTG
1851 ATGAGCTGT AGCTGTACT CAGGCCAGC AAGTAAAGA GGTATCCAG
1901 AAGACGTA AAGTGTGTC CCGTCTTCA ACTGTAAA ATTACACA
1951 GACCAAGAA AGAATTTGT GCTTACCGA AGAATAAT CTAAACATG
2001 AGAATATG GGAAGAAA TTGCAAAAT TAAATATA ATGCAAAAT
2051 CTAAATTA AAAAAAAA AAAAAAAA AAAAAAAA
2101 AAAAAAG
```

BLAST Results

Entry XSPOLYAB from database DMBL:
Human mRNA for polyA binding protein
Score = 3420, E = 0.0e+00, Identities = 1162/1243

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 707 bp to 1936 bp; peptide length: 410
Category: strong similarity to known protein
Classification: uncat
Prosite motifs: RNP_1 (10-18)
RNP_1 (112-120)

```
1 LHTDESGKSE GFCFVSEFEM EDAQKAVDCH NGKELNGFOI YVGAQKIVE
51 NQTELRTPE QHQQRITPY QVHLVYVNL DQGLDDELRK KAFSPETIT
101 SAKVHSGGR SGGFVCFES SPEEATVATP DQGRIVATK PLVVALAQS
151 EERQAVLTNE YHONASVRA YHQAAPPSS YHNTAVPQTQ NHAATFPSSQ
201 IARLRSPNM TAQGARHPF QHPEALJNG APRVPESTR PASSOVPMN
251 ETONVHST OTVCPRPAA AHAATPRAV YTPETETAC YHPCQHMA
301 QPQVTHOLA VHYQOETLT ASKLASAPQ EKQKLGELR FLIQNHPT
351 LAGITONLL EIONSELLH LESPELSRK VQKAVLQA QKAEATQSA
401 VNSATOVFFV
```

BLASTP hits

No BLASTP hits available

Alert: BLASTP hits for ORF2phes3_Em10, frame 2

PIR:DHUHA polyadenylate-binding protein - human, M = 1, Score = 1931,
P = 1.7e-199

PIR:I46718 poly(A) binding protein - mouse, M = 1, Score = 1928, P =
3.6e-199

>PIR:DHUHA polyadenylate-binding protein - human
Length = 633

HSPs:

Score = 1931 (289.7 bits), Expect = 1.7e-199, P = 1.7e-199

Identities = 384/415 (92%), Positives = 396/415 (95%)

Query: 1 LHTDESGKSGFVSEFEMEDAQKAVDCHNGKELNGFOIYVGAQKIVERQTELRTPE 40
Subject: 219 VHTDESGKSGFVSEFEMEDAQKAVDCHNGKELNGFOIYVGAQKIVERQTELRTPE 278

Query: 41 QHQQRITPYQVHLVYVNL DQGLDDELRK KAFSPETITSAKVHSGGRSGGFVCFES 120
Subject: 279 QHQQRITPYQVHLVYVNL DQGLDDELRK KAFSPETITSAKVHSGGRSGGFVCFES 338

Query: 121 SPEEATVATPDQGRIVATKPLVVALAQRKEERQALTYHONASVRAVPM-----Q 174
Subject: 339 SPEEATVATPDQGRIVATKPLVVALAQRKEERQALTYHONASVRAVPMVPMVPMV 398

Query: 175 RAPPSCGYHNTAVTQTHAATFPSSQARLRPSRMTAAGARPHPTQKPSAIRGAPV 234
Subject: 359 RAPPSCGYHNTAVTQTHAATFPSSQARLRPSRMTAAGARPHPTQKPSAIRGAPV 458

Query: 235 PFSTRPASSOVPMNHTQVAMTSTQVGPRAAAAAAATPAVTVPRYTAAGVPM 294
Subject: 459 PFSTRPASSOVPMNHTQVAMTSTQVGPRAAAAAAATPAVTVPRYTAAGVPM 517

Query: 295 QHNAQPVTHOQLAVHYQOETLTASKLASAPQEKQKLGELRFLIQNHPTLAGE 354
Subject: 518 QHNAQPVTHOQLAVHYQOETLTASKLASAPQEKQKLGELRFLIQNHPTLAGE 577

Query: 355 ITQKLLIDWSELLHLESPELSRKVQKAVLQAQKAEATQSAVNSATOVFFV 410
Subject: 578 ITQKLLIDWSELLHLESPELSRKVQKAVLQAQKAEATQSAVNSATOVFFV 633

Score = 315 (47.3 bits), Expect = 1.9e-27, P = 1.9e-27

WO 01/12659

PCT/IB00/01496

Identities = 71/163 (43%), Positives = 102/163 (62%)

Query: 1 LNTDESGKRGVGFVSTFHEDAQAVDDHNGKELHGRQITVGAQRKVERQTELKRTFE 60
+S G+ +V+P+ DA+VA+D H+ GE + + +Q B L+++
Sbjct: 130 VVCDHNG-SKGIGFVNFETQAAZRAIEDHNGHLLWOKVYVGRFSAKREAEALCARAK 188

Query: 61 QHQDITRYQVWLVYVNLGGIDDEALAKAFSPFGTITSKAVNM-EGGRSGGFGFVCF 119
+ + +K+ + +DEAL+ P + S KVN E G+SGGFGFV F
Sbjct: 189 EF-----THVTIRHFGEDHDEALDLPGP--ALSAVWVDESGSAGGFGFVCF 235

Query: 120 SPEEATKAVTDHGRIVATPLTVALAQREKQAVLTWYH 163
E+A KAV RNDG+ + S +V+ ADH DNG L ++Q
Sbjct: 236 EHEDAQAVDDHNGKELHGRQITVGAQRKVERQTELKARFED 279

Score = 214 (32.1 bits), Expect = 1.9e-14, P = 1.9e-14
Identities = 50/150 (33%), Positives = 87/150 (58%)

Query: 8 KSGGFGFVSTFHEDAQAVDDHNGKELHGRQITVGAQRKVERQTELKRTFEQKQORI 67
+S G+ +V+P+ DA+VA+D H+ GE + + +Q B L+++
Sbjct: 50 KSLGAYVNTQQPADAERALTQTHFVIGKGFVAINWSQ----RPSLAKS----- 96

Query: 68 TETQVWLVYVNLGGIDDEALAKAFSPFGTITSKAVNM-EGGRSGGFGFVCFSSPEEAT 127
V H+++KRLD ID++ L FS FG I S RV+ + SKG+GFV F + E A +
Sbjct: 97 ---GQGITIKLQHSIDAKALYTTFATGFIKCVQVCDNGSKGIGFVNFETQAAZRA 153

Query: 128 AVTDHGRIVATPLTVALAQREKQAVLTWYH 157
A+ H+G ++ + +V+ + +ER+A L
Sbjct: 154 AIEKHGMLLWOKVYVGRFSAKREAEAL 183

Score = 120 (18.0 bits), Expect = 4.8e-04, P = 4.8e-04
Identities = 30/99 (30%), Positives = 54/99 (54%)

Query: 70 YQVWLVYVNLGGIDDEALAKAFSPFGTITSKAVNM-EGGRSGGFGFVCFSSPEEAT 127
Y+ +LV +L + + L + PFP G I S +V H KS G+ +V F P +A +
Sbjct: 8 YPMASLIGDLHPDYEAHLEKTSFAGFILIRVCRDHITRSLGAYVNTQQPADAER 67

Query: 128 AVTDHGRIVATPLTVALAQREKQAVLTWYH 165
A+ H+ +K+ + +QR + + + +N + + + +
Sbjct: 68 ALPTHTDVIKGFVAINWSQDPLASQNGIFIKNI 106

Peptide information for frame 3

ORF from 45 bp to 707 bp; peptide length: 221
Category: strong similarity to known protein
Classification: unmet
Protein motif: NMF_1 (138-146)

1 RHPSTSTSTPT ASLYVGLHIF DYTEAMLYEK FSPAGPILSI RICROLITSG
31 SSNAYVYMO KTKADAGALD THNFDVIGKR PVKIMNSORD PLAKESQVGR
101 IYVHKLQHSI HNEALYQYS ATGFIKCVYVCDNGSKGIGFVNFETQAAZRA
151 AERAKRNGH MLNGKRVVY GQFSAKREAE AELGAAKRT PVVYKMFGE
201 DNDDELRLDL PAFGFPALSV N

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for OXFIptee3_fm10, frame 3

SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
PROTEIN 1) (PABP 1).. N = 1, Score = 1039, P = 5.7e-105

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1031, P =
4e-104

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1009,
P = 8.7e-102

>SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
PROTEIN 1) (PABP 1).
Length = 436

HSFs:

Score = 1039 (155.9 bits), Expect = 5.7e-105, P = 5.7e-105
Identities = 199/220 (90%), Positives = 205/220 (93%)

Query: 1 MNPETSPYFASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60
HNP8 P8YF ASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60
Sbjct: 1 MNPETSPYFASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60

Query: 61 HTKADADALOTNMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 120
DAE ALDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 120
Sbjct: 61 QPADALALOTNMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 120

Query: 121 AFQHLILCHVVCDESGKSGVGFVETHEAAEAALKHNGHLLMCRVYVGVQFSEKERE 180
AFQHLILC VVCEDESGKSGVGFVETHEAAEAALKHNGHLLMCRVYVGVQFSEKERE 180
Sbjct: 121 AFQHLILCHVVCDESGKSGVGFVETHEAAEAALKHNGHLLMCRVYVGVQFSEKERE 180

Query: 181 AELGAARAEFFWVYIKFPGEMDDEALDLFGTGFALAV 220
AELGAARAEF WYVYIKFPGEMDDEALDLFGTGFALAV 220
Sbjct: 181 AELGAARAEFFWVYIKFPGEMDDEALDLFGTGFALAV 220

Score = 275 (41.3 bits), Expect = 4.1e-23, P = 4.1e-23
Identities = 71/233 (30%), Positives = 120/233 (51%)

Query: 2 HNPETSPYFASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 61
HNP8 P8YF ASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 61
Sbjct: 90 DPLRLKSGVGMIFVEMLDKSIHKKALYDVSFAGPILSIRICRLITSGSNTATVWFQ 149

Query: 62 HTKADADALOTNMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 117
DAE ALDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 117
Sbjct: 150 AELGAARAEFFWVYIKFPGEMDDEALDLFGTGFALAV 209

Query: 118 TVBAPFHLISCHVVCDESGKSGVGFVETHEAAEAALKHNGHLLMCRVYVGVQFSEK 174
TVBAPFHLISCHVVCDESGKSGVGFVETHEAAEAALKHNGHLLMCRVYVGVQFSEK 174
Sbjct: 210 LFGKGFALSVKVFVETHEAAEAALKHNGHLLMCRVYVGVQFSEK 249

Query: 177 KEREALGAARAEFFWVYIKFPGEMDDEALDLFGTGFALAV 219
KEREALGAARAEF WYVYIKFPGEMDDEALDLFGTGFALAV 219
Sbjct: 270 VERQELKRLKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 322

Score = 227 (34.1 bits), Expect = 6.3e-18, P = 6.3e-18
Identities = 57/187 (30%), Positives = 101/187 (54%)

Query: 12 SLVYGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 71
HNP8 P8YF ASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 71
Sbjct: 192 WYVYIKFPGEMDDEALDLFGTGFALAV 250

Query: 72 HNPETSPYFASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60
HNP8 P8YF ASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60
Sbjct: 251 HNPETSPYFASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQ 60

Query: 115 LYDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 174
DAE ALDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 174
Sbjct: 310 LKEREALGAARAEFFWVYIKFPGEMDDEALDLFGTGFALAV 369

Query: 175 SKEREAL 183
SKEREAL 183
Sbjct: 370 KEREAL 378

Score = 100 (15.0 bits), Expect = 2.3e-02, P = 2.3e-02
Identities = 26/99 (26%), Positives = 53/99 (53%)

Query: 8 YPTASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 66
HNP8 P8YF ASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 66
Sbjct: 291 YPTASLVGDLKPOVTEAMLYKFSFAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 66

Query: 67 HALDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 106
DAE ALDTMDFVIGKGFVIMHSQDPSLRSGVGMIFVEMLDKSIHKKALYDVS 106
Sbjct: 348 KATVEMGRIVATPLVALAGPILSIRICRLITSGSNTATVWFQHTKADADALOT 386

Pendant information for DRFZptaa1_Bal0, frame 2

Report for DRFZptaa1_Bal0.2

[LENGTH] 409
[HW] 45235.68
[PI] 10.08
[HOMOL] SWISSPROT:PAR1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN 1) (PABP 1), O.0

[FUMCAT] 04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S. cerevisiae, YER165w] 1a-54
[FUMCAT] 30.03 organisation of cytoplasm [S. cerevisiae, YER165w] 1a-54
[FUMCAT] 30.10 nuclear organisation [S. cerevisiae, YER165w] 1a-54
[FUMCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YER165w] 1a-54
[FUMCAT] 04.05.99 other mRNA-transcription activities [S. cerevisiae, YER165w] 1a-15
[FUMCAT] 11.01 stress response [S. cerevisiae, YER165w] 1a-12
[FUMCAT] 04.01.04 RNA processing [S. cerevisiae, YER165w] 1a-12
[FUMCAT] 04.99 other transcription activities [S. cerevisiae, YER165w] 1a-09
[FUMCAT] 94 classification not yet clear-cut [S. cerevisiae, YER165w] 1a-08
[FUMCAT] 03.19 recombination and DNA repair [S. cerevisiae, YER165w] 1a-07
[FUMCAT] 03.13 mitosis [S. cerevisiae, YER165w] 1a-07
[FUMCAT] 04.05.03 mRNA processing (splicing) [S. cerevisiae, YER165w] 1a-07
[FUMCAT] 04.07 RNA transport [S. cerevisiae, YER165w] 1a-07
[FUMCAT] 30.13 organisation of chromosome structure [S. cerevisiae, YER165w] 1a-06
[FUMCAT] 99 unclassified proteins [S. cerevisiae, YER165w] 1a-06
[FUMCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YER165w] 1a-05
[FUMCAT] 08.01 nuclear transport [S. cerevisiae, YER165w] 1a-05
[FUMCAT] 11.04 DNA repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YER165w] 1a-05
[FUMCAT] 03.01 cell growth [S. cerevisiae, YER165w] 1a-04
[BLZKHS] HLO0308 Eukaryotic RNA-binding region RNP-1 proteins
[SCOP] dist_ 4.34.7.1.3 Sea-lethal protein [(Drosophila melanogaster)] 1a-17
[PIRSM] nucleus 0.0
[PIRSM] duplication 0.0
[PIRSM] RNA binding 0.0
[PIRSM] nucleolus 2a-09
[PIRSM] tandem repeat 2a-09
[PIRSM] single-stranded DNA binding 3a-06
[PIRSM] DNA binding 1a-13
[PIRSM] phosphoprotein 4a-10
[PIRSM] ribosome 3a-08
[PIRSM] mitochondrion 3a-08
[PIRSM] alternative splicing 9a-11
[PIRSM] chloroplast 2a-19
[PIRSM] transcription regulation 2a-07
[PIRSM] protein biosynthesis 3a-08
[SUPFAM] nucleolin 6a-10
[SUPFAM] glycine-rich RNA-binding protein 2a-07
[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 2a-19
[SUPFAM] polyadenylate-binding protein 5.0
[SUPFAM] ribonucleoprotein repeat homology 0.0
[PROSITE] RNP_1 2
[PFAM] RNA recognition motif. (aka RRM, ABD, or RNP domain)
[SWI] 1 repeat
[SWI] 3D
[SWI] LOW_COMPLEXITY 5.42 1

SEQ HTDSEKSGGPGFVPERHEDQKAVDHWGHELMKQIYVGRAGHYVERTELRTTTEQ
SEG
1a1-
SEQ HQQRITATYQVWLVKVLGDIODEALKAFTFPGTITSARVYHGGHSGPGFVCFES
SEG
1a1-CEEECCCTTTTNNNNNNHTTTTCCCCCECTTCTTTEEECTTT
SEQ FEETATVTDHGRIVATKPLVALAQKKEEQALTYNQHNASVNAVHQAAPSGY
SEG
1a1-HHHHHHHHHTTTTCCCCCCKCKC
SEQ FWTAVPTQHHALYFPSSQIALRLSPFMTAGGARHPFQWPSAIRPCAPVFTNRP
SEG
1a1-
SEQ ASSQVPHVSTQVARTSTQVGPFAAAAAAATPAVTVPRYTAAGVHVPQRRMAQ
SEG
1a1-
SEQ PVTNODLAVHYGGDTLTASRLASAPFQKHGHEALFPLIQNHPTLAKITGKLE
SEG
1a1-
SEQ IDMSLLTLESPEGLASVDEAVVLAQQAQKATQAVHSATGVTV
SEG
1a1-

985

[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 3e-25
[SUPFAM] polyadenylate-binding protein 1e-112
[SUPFAM] ribonucleoprotein repeat homology 1e-112
[PROSITE] RNP_1 1
[PROSITE] RNA recognition motif. (aka RNM, RBD, or RNP domain)
[RM] All_Rna
[RM] 3D

SEQ ERSLVCLRAAVFHWPSFSTPTASLVGDLHPNTKAMLYETSPAGPILSICRDL
[ha]-EEETTTTTTCHHHHHNGGCCCEEESEET
SEQ ITGSSHYAVVHTKDAHALOTWTFVIGKFWIRHNSORDSLKSGVNI PVKL
[ha]- TTTCEEEEEEKCKKHGHHKHHTTEE--TT---EEEEEETTTTCCCEEESEEC
SEQ DRSIHWALDTVAFGWILSCVYVCDKMSGYGVYFETKLAARAKKHQKLLGR
[ha]- TTTCHHHHHHHNGGCCCEEESEETTTTCEEESECKHHHHHH
SEQ KVPYGOVSRACEAEAGARAEFFHWYTKMTCEDHDEKLDLPKPGPALSVH
[ha]-

Prosite for DKF5piles_0010.3

PS00030 152->160 RNP_1 P0000030

Plan for DKF5piles_0010.3

NAME	RNA recognition motif. (aka RNM, RBD, or RNP domain)
HEM	*ITVGLPWDTTEEDL-DIFQFQPLVHIMHQRATGASGFAVFEED
Query	27 *TVGLL+DTE+L+PS+QPI+SIR+RD+Y+S+AV+V+*
HEM	EEDEALIdenQmFWCRIRV*
Query	76 DAE A+D+RH ++ G+++R+ 98
HEM	*ITVGLPWDTTEEDL-DIFQFQPLVHIMHQRATGASGFAVFEED
Query	115 I+V+GL+...+L+D+S+EQ+IS+...+D++S+G+...+V+PS+
HEM	EEDEALIdenQmFWCRIRV*
Query	162 *EAE+AI+HMGH+++GR++V 184

WO 01/12659

PCT/IB00/01496

DKFZphtes3_bf7

group: testes derived

DKFZphtes3_bf7 encodes a novel 412 amino acid protein without similarity to known proteins.
No informative BLAST results: No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits (both from testis libraries)

Sequenced by MediGenomix

Locus: unknown

Insect length: 2859 bp

Poly A stretch at pos. 2870, polyadenylation signal at pos. 2852

```
1 CGAGACGGCC CTTGGGTGCT GGTGGCGTGC CTCTGCTGGG CCTGAGGAAA
51 ACACCTGCCA TGGCGCAGGG CGGGAGCGCC GAGGAGGCC CCCACTCGGC
101 CGGCGAGGCG TCTCTGTGCG TGAGTGGGCT GCAGGAGTTC CTTAGAGAGA
151 ATTCTATTTC CTACACGACA ACACCAATTT CTACCTTTGT GGAATTTATG
201 TAATATTATY TAATATTGAA ACCAGAGAAA AGACTGTACT GCACCTGTATG
251 AATGGAATTC TGGGGCTCAT GGCACCTTAC ATGCTCTCTG AACTTTGGGG
301 TTTTCTTCAG CGGAGGCTAA AACCTCTCAT CTACGTATAC AGCTTTCCAG
351 GATTGACAG AAGGACCAAA TTGAAAGGCA ACATTCTCTT GGACTTACCT
401 TTACTTTTAT TCACTTACTG TGGCACTTAC CTGGCTAATY ACTCTCTCTY
451 CCCAGATTTT GAATGGGCC TTTGGAACCT GGAATCAGAT ATCATTTTGT
501 GTAGAGATTC ACAGCTCTGA ATGAGTGTGA AGCAATCTTC TTTTACCCCG
551 ATGAACTTGG GCGAGCTGTG CTATATCAAT CCAGATACAG TGAGGTTGTG
601 GACATTTGGA AGAATTAACC AGAGAGATTC TTTGAGAGA AGTGTGTTGA
651 AATTACCTCT AGAGATGGG TCATTTTATA ATGAAGGGA TGTCTTTTC
701 CCCCACTCTG TGGCGAAGGA TCTCATCTAT AGTCTCTG TGACATCTCT
751 AGCAATCTCC GGTCTGTGAG CGAAGAGGCT AGCACTTTTC CGGCGAAG
801 ATGATATATA TCTTTGCTTT CAGCGGACTA TGCATTCTGT GACTCCACAA
851 ACTGATCTTC ACATTGGCTG TGAGAGGGGT CATCTTTTAA TGAATTTG
901 AGACACTCTG CAAGTGACTG TACTTAATAA GATAGAGAG GAATCCCAT
951 TGAGAGAGAG AGCAATTTT ATGAGTCTGA TACCTTTGTT ATATCAGAG
1001 GAGGGGCTGC TGGCTTCTGG AATTGATGGC TTTGTGATTT GTTTTATAT
1051 TAAGATGAGA AGTTACATGA TGGAGGATTT TCTTGAGATT GAAGACCTTG
1101 TGAACATATY GACAATTTCT CCAATTTTCT CACTTTTCTY GATTCAGAGA
1151 GACAGGGGAT CTGTTTATAT CTACACTTTC GGTAGAGAG CAACTTTAAA
1201 TAAGCTCTTA GATGCTGTG ATGGGAATTT TCAAGCAATY GACTTTATCA
1251 CACCTGGAAC CGATACTTCT ATGACACTTA CATATTCAG GGAATTTGT
1301 GTTTGTTGGG TGAGAGATTC TCTTTGTTA AGCAAGATTT ATCTGAATAC
1351 CTAGAGAGAG GTTCTGGCTT GCTGTCTCTC CTGCTCTCTY GCAGCCGTGG
1401 GCAGGAGAGA TGGCTGGCTC TACTTCTATC GGTATATAGA TANGAATTC
1451 CTTGAGTCTG TGACAGAGGC CTCTCTCTCT GAATCTCTCT TGCAGACCT
1501 GGTGTAAGTC CTCTCTGCTT CGAGAGCGCG CTGCTGTCTA CACCTCTCTG
1551 TTGAAGATTC TATGAGAGCT ATCTCTCTCT TTAATTTTAA CTTTAGTGT
1601 TTTTATTGTT TTTGAATCTT AATATATCCA CACACTTTAA CACTCAAAAG
1651 GTACAGAGGG CTGTTAGTAT AAGTACCGCC CATAGCCAGG TCTCTCTG
1701 CAGGAGGCTT GGTACCAATY TCTCATCTCT CTCTTGAGAT GTTTATCCA
1751 TGACAGAGCA AATCATATAA AGCACTCTTT TTTACTTTTA TCAATGGCCA
1801 TGATGTTGCT ATAGTGTGCC AGCACTCTTT GCTGTATTTA CTCTATGAG
1851 TAAGACCTCT TGTGTCTCTT ATTTGACAGG TGAGGAGAT AAGGACAG
1901 GATTTTAATY AACTTTCTCA ATCTACACA GATATGAAI GCAATATCT
1951 GGAATTTGAA CCGAGGTAGT TGGGCTGAGC AGTCACTGCC TTTGCTCTTA
2001 AAGAGAGAAA ACATATGACA ATGCTCTATT TCTTTTCTCA CTTAATCTTA
2051 TATCTTGGAG ATGTTTATAT ATCCACATAT AAGAGAGAG CTAATATTT
2101 GTATAGCCAG ATAGATATCC ATTATATGAA TATACATCA TTTTATAAA
2151 ACCTATATY AATGACATAT TGAATATTTT CAATACTTTT GAGGATATC
2201 TTTTAGATG ATATATATGA GACATTAGAT TTGAGCTTGT AGGTCTATC
2251 ATATATCTG TTCTTTTATA ATTTATTTTA TTATTAGTAT TTATAGAGAA
2301 CAGCATTTTC TATCTGCTTT TACACTTTGA GATCACTTTA GCTCTTGGCA
2351 TGTATCTCTC AAACACCAAG TCAGAAAGCT GTTATCTTTA TCTCTATTAG
2401 ACAAATTAAG GAATTTAGGG TTGAGAGGCT GAGGAAAGC ATTTCTCAG
2451 ATTACACTT ACAGAGTAG CACACTGAGG AGCTGGCCTT GGCACCTGG
2501 ACTGCGCAGC TGCACAGGCC TACCTCACTG GGAAGAGATG GATATCTCT
2551 TTTCAATTTT CCTCTGCTCT TCTGACATCT CATTTTGTGT TGCCTTTCT
2601 TCTCAGATCT CTCTATCTCT AATTACATCT TTCCCACTTT TCTCAATTTG
2651 ATAGATGTT AGACATGTTT CACTTCACTT TCTCTGAGC TGCAGGTC
2701 CAGACAGGG GTAATGAAT GTACACCCA CCACATATTT GAGATTTGCT
```

WO 01/12659

PCT/IB00/01496

2751 TATTTTCGGT TGAACATCA AGAAGCTCT ACCGACGAC ATGTTTCATT
2801 CACTATGAT GAGCAACCTG CCAATCTTAA CTGAATCTTC TTGACTGTAT
2851 TATTAAAGT TGCATCTTGC AATTAAGAAA AAAAAAAGG AAAAAAGG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 269 bp to 1504 bp; peptide length: 412
Category: putative protein
Classification: no clue

I HATHIACEVW AFSDEKLEPL IYVTSFGLT RATAKGNIL LQVLLSFRY
51 GQVTLARYS LPETELALRW HESITLCSA QDQNDVQNH SPANWVQRL
101 CLSPSTVSV WTIRSNQEN CFKASVLEP LEDGSFNET DVVFPQSLPK
151 DLVGPVPLP SAJAGLVKE AETFRPDGL TPLAFTPHC WYTFSLYIG
201 CSDHLNLMH QDTQVTVLM FIEESPLED RMTFSMTL VYQKCVLAS
251 QIQPVYSPI IKDSIHIED FLIEFPIER HTSPWTVLL LIQDQGVYV
301 IYTFGEPTL RYVLACAGCA FVLSIFITPC TQVNTLTVS QCLVWLELD
351 CACVSKITLM TLATVLACCP SLSAAVOTE QGVYFISVT QRESQVVRH
401 AFLSESVQHV VV

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF5phtes3_Rp7, frame 2

No Alert BLASTP hits found

Peptide information for DKF5phtes3_Rp7, frame 2

Report for DKF5phtes3_Rp7.2

[LENGTH] 412
[MW] 4476.62
[pI] 4.91
[FW] Alpha_Beta

SEQ HATHIACEVW AFSDEKLEPL IYVTSFGLT RATAKGNIL LQVLLSFRY GQVTLARYS
FRD ccc

SEQ LPETELALRW HESITLCSA QDQNDVQNH SPANWVQRL CLSPSTVSV WTIRSNQEN
FRD cchhh

SEQ CFKASVLEP LEDGSFNET DVVFPQSLPK DLVGPVPLP SAJAGLVKE AETFRPDGL
FRD hhh

SEQ TPLAFTPHC WYTFSLYIG CSDHLNLMH QDTQVTVLM FIEESPLED RMTFSMTL
FRD ccc

SEQ VYQKCVLAS QIQPVYSPI IKDSIHIED FLIEFPIER HTSPWTVLL LIQDQGVYV
FRD ecc

SEQ IYTFGEPTL RYVLACAGCA FVLSIFITPC TQVNTLTVS QCLVWLELD CACVSKITLM
FRD ecc

SEQ TLATVLACCP SLSAAVOTE QGVYFISVT QRESQVVRH AFLSESVQHV
FRD hhh

(No Prosite data available for DKF5phtes3_Rp7.2)
(No Pfam data available for DKF5phtes3_Rp7.2)

WO 01/12659

PCT/IB00/01496

Defigntes1_9e22

group: testes derived

Defigntes1_9e22 encodes a novel 227 amino acid protein with weak partial similarity to Ring-finger proteins.

For the novel protein, Pfam, but not Prosite predicts a C3HC4 type RING finger motif.
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to zinc finger proteins

Sequenced by DMFS

Locus: unknown

Insert length: 1318 bp

Poly A stretch at pos. 1308, no polyadenylation signal found

```
1  CCTCCGCCCG  CTTTCGGAGC  CCGGGGGGGG  CTTGTGGGGG  GGGGAGCCCG
51  GCGCGGACTG  GCGCTCTTTC  GACCTTGAGG  GGAACATGCG  GTTTCCTTTC
101  GATCGTTTGA  AATCTAATGT  TTGGATGCG  GCGCGGGGCG  CTTGACTCTT
151  CCGCGGCCCG  GGTTTTTCGC  TTTTTCCTTT  TTGCTTTTTT  TCTTTTCTTC
201  CCTCCGGGTC  TCTTTTTCGA  CTCCTTCGCC  CTTTATGCTC  GCGCAGGCTC
251  CCGCTCTCTC  CTGAGAAATG  GGGGAGGGTC  TGGGCTCTCA  GGTTCGGCGC
301  CCACCGGGGC  CCGGGCGAGC  ATGGGGGCGA  AGCAGAGCAC  GCGGGCGGGC
351  TCCCGGAGTC  CTTTCGGGG  GGTTCGACAC  GATACAGAGC  CGTTCGGGCG
401  GCGGGGAGGG  GCGGCCCAT  TCGGGCACTA  CCGGAGGGGC  GCGGGGGCCA
451  TGGGGTGGG  CAGCGGCTCG  CTCAGCTGGG  TGGCAGGCAT  GGCATAGGAC
501  CCGCAGAGGG  CCGGGAGGCT  GCTCTTTCGG  CTCAGAGGCC  CGGCTTCGCG
551  GCGCAGCGCG  GACTCCGAGA  GGGGGGCGCG  GGGCGAGGGG  TCTGCTTCGC
601  ACTGACACTA  TGGCATGGC  AATGCTTACG  AGGAGAGGGG  CGGCGGTAC
651  CATAGAGAGC  GGATGCTGTA  CTTGGGCTTC  CGAGCTTCGC  TGGCGGATGC
701  TCTACTCTCT  CACATCGCAC  CGAGGTGGTT  CAGCTCTCAT  AGTGCTTCA
751  AGTCCCAT  TTGCTCGAG  TCTTGGCTTT  CTGACGAT  GGAATGAC
801  TTATATATGT  GTTTCAGCAA  ACTCGGCTTC  TCTACAGAG  ATGATGTGCT
851  GACTAAGAGC  GCGGTGAGT  GTTGACTCTG  CTTGAGAGAG  CTTCTGAGG
901  GGGACAGCAT  AGCCAGGCTC  CCTCGCTCT  GCATCTATCA  CAAGAAGCTC
951  ATAGACTGCT  GTTTTGAAGT  GACAGATCT  TCTCGAGAC  ACCCTGGGA
1001  CTACTCTCTG  GCTTCTTTC  CTGACTCTTC  TGAAGGAGC  AGAGGCGCGC
1051  TGCTTCAGGG  AGAGGCTCA  CCGGACCTTC  GGGCAGAGT  GAGCTTGGGA
1101  CACAGGGGG  AKAAGGAC  CCTTCTTCT  CTGACTCTCA  GATATGTT
1151  CTCTCTCTCT  CCTCGAGAG  ACCAAATGCG  ATGAGAGCAA  GTTTGAGAGA
1201  AGATGATATC  ACTGCTATC  CTTGCTCTCA  CCGCTGAGCT  CAGAGGGGAA
1251  AGGGCATTTT  CTTTTCATC  TTTGAAGGC  ATTGTGGTC  TGTCTTAA
1301  GTTTTACAA  AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 321 bp to 1001 bp; peptide length: 227
Category: similarity to known protein
Classification: unclassified

```
1  NGGKSTAAK  SAGPFFVST  DGSAYFFPG  AFHFGHYATC  GGAGLKRKS
51  VSVAGMGHD  PSTAGGVFG  LTPASRGTG  DSRAPGGG  SASDITYANG
101  NGVGTGGD  WPDGVLGCS  RAGLADALPL  HIRPNVSH  SPTKVLCS
151  SVASDEMDM  FIMLEKPL  SYRWDLTED  AGEDVICLES  LIGQITLRL
```

WO 01/12659

PCT/IB00/01196

Z01 PCLCIYHNSC IDSWEVHNS CPEKPAD

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRF2phtes3_9e22, frame 3

TRZMBL:AF078822.1 product: "RING-H2 finger protein RHA2b"; Arabidopsis thaliana RING-H2 finger protein RHA2b mRNA, complete cds., M = 1, Score = 111, P = 2.8e-06

TRZMBL:AF078822.1 product: "RING-H2 finger protein RHA2a"; Arabidopsis thaliana RING-H2 finger protein RHA2a mRNA, complete cds., M = 1, Score = 112, P = 6.4e-06

TRZMBL:AC004131.14 gene: "T17M13.17"; Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence, complete sequence., M = 2, Score = 123, P = 1.4e-05

PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana, M = 1, Score = 142, P = 8.8e-08

>PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana
Length = 327

NSFs:

Score = 142 (21.3 bits), Expect = 8.8e-08, P = 8.8e-08
Identities = 24/57 (42%), Positives = 30/57 (52%)

Query: 166 SLPRLSYNDVLTGDAECVLCLELLGGDTIARLPLCLIVHNSCIDSWFEVHNSCP 222
S P + L P D + C + C I E E + G L P C 1708 C I + W S C P
Sbjct: 206 SLPVVKITPQLTHNSOCTVCHLEFIYGGDTLPCHIVHNSCIVPLALHNSCP 262

Pident information for DRF2phtes3_9e22, frame 3

Report for DRF2phtes3_9e22.3

(LENGTH) 227
(MW) 23782.62
(pI) 6.18
(HOMOL) PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana 2e-08
(FUNCTION) 19 unclassified proteins [S. cerevisiae, YDR313c] 4e-06
(FUNCTION) 30.07 organization of endoplasmic reticulum [S. cerevisiae, YOL013c]
(FUNCTION) 0.001
(FUNCTION) 08.13 proteolysis [S. cerevisiae, YOL013c] 0.001
(PFAM) Einc finger, C3HC4 type (RING finger)
(TM) Irregular

SEQ HGGQSTAAASRGPFPGVSTGDAVFPFGGAPHTGRTYGGAGLANSVSVAGQND
PRD ccc
SEQ PSTAGGVFFGLTFPASNGTGDERAPOGGGASDSTIANGGYOETGGHHRDGNLILCS
PRD ccc
SEQ RASLADALPLHIAPIPMFSHSGFPCPCISASVADSDNDFTNCLSPRLSYNDVLTED
PRD HNNNNNNHNN
SEQ AGCVLCLELLGGDTIARLPLCLIVHNSCIDSWFEVHNSCPKPAD
PRD ccc

(No Prosite data available for DRF2phtes3_9e22.3)

Pfam for DRF2phtes3_9e22.3

NAME Einc finger, C3HC4 type (RING finger)
HMM *CPICFCTQIDyWFFPdeWmLCkHsPcypCfzW.....CfmC*
C IC L+++ D++ LPC+ ++ +CI +W CP+
Query 184 CVIC-----LEELLGGDTIARLPLCLIVHNSCIDSWFEVHNSCPEN 224

WO 01/12659

PCT/IB00/01496

DKF8phtes3_9120

group: testis derived

DKF8phtes3_9120 encodes a novel 205 amino acid protein with similarity to human KIAA0336 gene.

No informative BLAST results: No predictive prosite, piam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKF8

Locus: /map="44.1 cR from top of Chr17 linkage group"

Insert length: 2509 bp

Poly A stretch at pos. 2499, polyadenylation signal at pos. 2481

```
1 CTGGCCGAGA TGACCTGGGC ACCTCTGGCT TGAATGGGCA AATACTGATC
51 AAGCGCCAPT TATTCTCTCT TCAAGAACTC TAAGCTTAGC AGAGAGAGAT
101 AGGCGGAGAG AGTTCATCAA TGCGCTGGCT GGAGGACAGC CAATTCAGG
151 ACATTGGCAA CGGAGTGATC AAATCATAG ATCATGAGGC CTAAATGAA
201 TANGGAAGAA AGAGAAATGG CAGAGGCTGA GAACAGAAAG AGAGGCTGGA
251 GGGGCTCTTA ATCTTGAAIA TTAGGCTATA ATATAGCTAT ATGGGTAAAG
301 ATTGAGAGAA TTGTGTAGGA GGAAGTAGTC AAAAAGTAGA AGCAGTTGG
351 AGAGTAGATT ACAATATACA AGAGGAGGCT GGTAAAGAG TGGAGCTATA
401 GGTCAATTGA GCTCAAGAAA CTGAGTCTCT AGGCGATTGG TTAGTCTATC
451 TGTCTAGACT TCAGAGTTGT CTAGAGTGA AATTCAGAG ACTGATCTGT
501 GCGAAGCTCA CAGTTTCTTC AGCACTCAAA ACAACATAGC AATATAGCC
551 AAGATCTCTG TGGATCAAT GACCTACGAG GCCCGTTCT TTGCTCTCAC
601 GCGCAAGAGC TGATCTGCTC GAATCTCACT TGAATTTCAA GACTACCTAT
651 TTGAAGTAGT CGAGGCGCTT GACACAGGTA TTCTGAAGAA GCTGAGTGG
701 TTCLAGACT GTGACATTAG CCGAGTTCAG ATTGCAAAAT GCACAGAGAA
751 GTTCTCTTTC TTGATGAAG GACATTTTGA TAACTTTTT AGCAAAATGG
801 AGCAACTGTT TTGCAAGCTG ATTTTACGTA TTCCCTCAAA CATCTCTGCT
851 CTTGAGATA AATCTAGGCA GCACTCTTAT AGTGAAGAG ATTCTAGCA
901 TTCTCAGAAA GAAATTGAAC AGTTACAGCA GAATTCACAG ACTGATTTAT
951 CTACTTAGCA GGGGCTCTCT GCGAGATTAG AGAGGAGAAA AATPTCTAG
1001 GCGAAGCTCA AAGAGAGGCT GACTTCTCTT GATGAGCTTC ATATCTTGG
1051 CAGAGATCAT GGGAGTAGTG ATTTTAGGGA GAGTTTAGTA TCCCTGGTTC
1101 AGAATCTCAG AAATCTCAG ACATATACAG ACATATGAGA AAGGAATCTG
1151 AAGGAGCTGA AATATCTTTA ATTGCTCAGT AGTCAAAAGC AGGAGGCTGT
1201 CAANAGTAG ATCATTAAGG ACTGTTCAAA CATTAGAGAC TGTCAAACTC
1251 ATACAGCTGA CTCTCAAC CAACCATACT TTTTATTAGA TTGCTCTTGT
1301 CACTCTCTTC TTGATCTGCT TGTCTCTCTC TTTTGTGATC CACTCTCTGT
1351 AGGATATGAG TGTACTACTT TGAATAGGC TGAAGCATCT GAGTCTCTTA
1401 ATAGTGGGA AGGATGCAAA CAAGAGAGCC ATGACCATGT AAGATATTTT
1451 GCGAGATCAC ACCTTGCTGA TAACTCTTCT CTGACCTTGA TTATTTTGGC
1501 TTACTCTTTC GATGAGGACA GACAGAGAAA GGATTAAGCC GTTGGCTCTCT
1551 TTATATTAT TATTATTGTT TTGAGAGACA GTTCTCTCTC TGTACAGGAG
1601 GTTAGAGTAG ATTTCAGTGG CACAACTCTG GCTCACTGCA ACCTCTGTGT
1651 CTTAGGCTCA AGTATCTCTC CTGCTCTCAG CTGCGAATA GCTAGAGACA
1701 CAGCTCTGCT TCACTACTCT TGGTAAATT TTCTGAGAA ACAGAGGCTC
1751 ACTATATTGT CAGGCTGAG TGCTCTTTTT ATTAACAGAT CATTACAGCT
1801 GCGAGAGGAG AAGATAGAGT AGTGTCTCTC GTCTGTGAAA TTCTCTTCA
1851 TGAGGAGGCT AATATGAGT GGAAGAGAGC ATGTAGCAAA AAGAGCAACC
1901 TTACTCTTCA ATAAAGAGAG AGTTGTTTCA TTGCAAAAT AATCTCTGCT
1951 ACAAAGAGCC TGATGATCTT AAGCAATTGA CTGCTTTAGA GTTCAGAGAA
2001 AGACCTTAGA CAAAGAGAGC AGACCTCACT GAGTAGAGAA AGGAGCATGT
2051 TAGATATAG TCACTGTTGA AATTAATT TACTACTGT TAGATATCTA
2101 TTGCAATTGT TTTCATACT TCACTGTGTT TTGAAATCTG CCTATACCTC
2151 AATATTTAT TATTCAGACT CATAGAGATC AATATTTTAA TGGCTCTAGT
2201 GCGAAATTTG TGTTAAACT CAATGGAATC TAATATTCTT TTATCTCTGT
2251 ACTGCTCTGA AATCTTAGG TCACTAGAG AAGGAGAGAA AATAGCATGT
2301 CTCTCTCTGA AGATATCTCT TGGCTCTGAT GTTCTCTTAA AGAGAGAGAG
2351 TTGAGGTTTC TCTTTATGT AGAGAGAGAG TAACCTAGGG TGATTTTGA
2401 ATGAATATT CATGATATT GAAGCTCTCT CTCTACATGA AATATCTTAA
2451 TTATCAAGAA GTCTTTTTTC CAATCTCTGA CATTAAATAT ATGTCTTTTA
2501 AAAAAAAAA
```

BLAST Results

PCT/IB00/01496

NO Medline entry

NO BLASTP hits available

KSPa :

Pedant information for DKFZphtes1_9120, frame 2

Report for DRYphtes3_9120.2

992

PCT/IB00/01496

(No Prosite data available for DKFZphtes3_9120.2)
(No Pfam data available for DKFZphtes3_9120.2)

WO 01/2659

PCT/IB00/01496

ORFspntes3_9k22

group: testes derived

ORFspntes3_9k22 encodes a novel 304 amino acid protein with partial similarity to X. laevis satenin p80.

No informative BLAST results: No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C-terminus of satenin p80

Sequenced by ORF3

Locus: unknown

Insert length: 2678 bp
Poly A stretch at pos. 2665, no polyadenylation signal found

```
1 CTCTCTAGGC TCCCGGGGCG TGGTCTCAG GCGCGAGGCT GCGCTGAGGC
31 GCGCGAGTAC CATTAGGCGC CGGTACTTAA GAGATTATGC CATCAGAAAC
101 CCACAAATTT AAAAACCAGA ACTTTTGTAA TAGATTATGC GATCAATTCA
151 TTGATCTTCC TAGAAAAAGG ATCTCTAATT TCACTATAAA GAACATGAG
201 GAGCTTANGA AATCTCTGAA ACCTCTGGCT GCTTCAATAA ATAGAACACT
251 TGGCAAAACT CTGAAAGGCG CAGATAAACT TGTAAAGTGT ATCTATGCGA
301 GAAGAAAGAT TGTATCTGCG TTCTCAATCT CTCTCTACAG AAAAAGACAG
351 TCGCTCTGAA GTGGGGGCTG TGACATGGCA ATAAAGAAA ATGACATGGC
401 TTGTGAGGCG CAGCTCTCTG AAAAATGACA CAGAGAGT GAAACATATT
451 TGGTAACTGC CAGTGAATCT GGTCTCTCAC AGACAGAAAG CCAATCATCA
501 AAATATATGT GCTTTTCTTC TGAGTTTCT CAGGACATG AAACATGGCG
551 CGAGCTTCTG TCGAGAGGGA ATATGAGATT GAACTAGCT TTAACTTCT
601 GAGAGAGAGG AAGTATAAGT GAACCTGTAG CTATTCTTCT GAGGATAGAA
651 GATCTTGGCG TTCTGTTGCA TTGCTTCTCT CTCTCTACCA ATTCTTACA
701 GGAAGAAAAA CAAATATATCT CACTTCTGCT CTGTCTGAC TTGTCTGCTC
751 TACTTAAGTC ACTATCTTAA AGCAAAATTC AGAATTTCT TATGTGCTG
801 TTAACCTGGC TTCAAGCACT CATTAAGAG TGGTCTGAG AACATCATC
851 CAAGACAGA ATTAATATG ATGGAATAT TAAATTTTA AAACACAT
901 TAACTGGATT ATGGAGACAG GAAACCATC TTACTTTGCT TCGAGATAT
951 ACTGATATA TAGTANGA TGTAGATCT TATTTATAC AGTTAGATTT
1001 AGAGATCTCA TTACTAAGG AGCAATTCCT TTTCGAAC AGCTCTGAC
1051 TGTATAAATT ACAAAGAAA AGCTCTGCT TGAGAACTCT GAACCTGTGA
1101 AGAATTCGAA ACTATTTTCT TTCTTAAGA CCGAGTAAT GAAGCACTA
1151 ATGAATCTCC AGCAATCTGC TTCAATCTCA ACTGAAAAA TATCGAAGG
1201 GAGGAGCTTC AATTCTATTC AGGTGAAGT GCACTATGAA GATTTCTAC
1251 CTTTCTGCTA TTGCGAGTAT ATATGTTAT TTGTACATAT TAGGAACCTAC
1301 TGAATTTTAA TGCATCTCT CATTAAGATA TATTTATAC TATCTGAAA
1351 AATAGACAGC GACTTCTCAC TAGAAGCAC CAGAGCAAT CATCTATAC
1401 TTTTATTAGA TTCTGTTTAA TTAAGAAAA AAACACCTA AATGTTGCA
1451 GCTATTTTCT TATCTGTGAA AGACTGAGG TTAAAGCAT GAAGAAATC
1501 AATATTAAAC ATTTTCTTCT CACATGAGA TACTGTATAT GTAAATCTCC
1551 TTAATTATTA ATAGGCTAT GTTGTGTGAT AGCAATATCT GTTTAABAA
1601 ACTAAACACA ACCATCTCTC TGCACTGATA AATCATGGA ATTAATATC
1651 GGGTTTACAT TCTGTAGG TGTCTGTGA ACATCTCTC CACATTTT
1701 AAACTTGA AATATCTTTA GTATCTCTCA TATTTCTGC CAGAACTATC
1751 ATCTCATGTA TGATGTGTT ATCCCTATCT AAGGAAAGG GTGATATGCT
1801 TTTCTTGA ATGTTTAACT GGAATCTCT ATGACTTGG CTAAATTATA
1851 TTACTTTT ATTTGACATA GATTTTAAAT ATTTCTAT CTCTGATCAT
1901 TTAACCTTCC TCTATCTGAG TAAATCTCAT AATATTTCT ATTTTCTTC
1951 TTTTATAAT TCTGATTTTA TATGAATCT AATCTTTT CACTGATAT
2001 GTTTAANGA GTTACTACA GTATTTAGA ATGTTTACA GTTAATCTG
2051 ATCTGTAT TTAAATCTCA ACATTTTGT TGACTATCT CTCTAATGCT
2101 TATATATA TCTACAGCA CTCTATGAG TCTTTTATA AATATGACT
2151 TTATGCTCA GTAAACAAA TTCTGATCA TTCTAGAGC TCTAATCTT
2201 ACTTCTCTG AAAATGATT GGTTTTATA GCACTCTG TGTTTCTAC
2251 ATCATCTCT CTTATATG CTCTAGAGG ATTTGAGAA AGTFTTTAG
2301 TAATTGAAT TCTCTGAGT AGCTCTCAA GCATGTGAT GTAAAGCTT
2351 AGAATTTAT CTTTATGCA CTACTATG GATCTGAG CAGATTTCT
2401 GCTCTTAAA GTTATCTTCT ACCTGATCT GGTFTGTCT AGCTATATT
2451 TATATATA GATATCTCT ACATGAGA TGAATTTTA TATTTGTTA
2501 TTCTCTTAAA TCTTATGAT ATAAATTT CAAATATT CACTCTGCT
2551 ATTTGAGG TGATTTATT AAGAGGCTAG TTTCAGGAC ATGGAATTT
2601 GAACTTAA CTTTATCT CTCTGAGT TTTCAGTCA TAAATATTA
2651 CTTTCTTGT TCTTAAAAA AAAAAA
```


WO 01/12659

PCT/IB00/01496

BLAST Results

Entry 09341356 from database DDBL:
human 878 nt-11840.
Score = 1267, E = 7.1e-50, Identities = 271/281

Medline entries

78227670:
Katanin, a microtubule-severing protein, is a novel AAA ATPase
that targets to the centrosome using a WD40-containing subunit.

Peptide information for frame 3

ORF from 87 bp to 998 bp; peptide length: 304
Category: similarity to known protein
Classification: unclassified

1 MARETHAVKRW KMKCKRDEH FIELEKXIS RFTVSRKCV KESPOLAAV
51 IRETHGQVYS SIKELKSTII KALEKSHFFP RCTKREKQD CGZGDMARE
101 EHELACAGHL KELLANDHPT IYVSSDSCS RQTERPRRY RQTFREYQD
151 IETNADVLPS RHHSLPVALY PPKESISEL VAILAIEDL QVVDCLPLH
201 YWLGKREKQY IALGCVOLL PLVKELLES FEETVIVLML VLQAVIKRW
251 HELSRTEIT KDMIDILAD QLSGLKQEN HLIVPDTTG WIAFDVDAIS
301 LQW

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DFFPhtes3_9k22, frame 3

TREMBL:AF056021.1 product: "p80 katanin"; *Xenopus laevis* p80 katanin
mRNA, partial cds.. E = 1, Score = 146, P = 1.2e-07

TREMBL:AF052432.1 product: "katanin p80 subunit"; *Homo sapiens* katanin
p80 subunit mRNA, complete cds.. E = 1, Score = 150, P = 1.2e-07

TREMBL:AF052433.1 product: "katanin p80 subunit"; *Strongylocentrotus*
purpuratus katanin p80 subunit mRNA, complete cds.. E = 2, Score = 146,
P = 1.2e-07

>TREMBL:AF052432.1 product: "katanin p80 subunit"; *Homo sapiens* p80
subunit mRNA, complete cds.
Length = 655

HSPs:

Score = 150 (22.5 bits), Expect = 1.2e-07, P = 1.2e-07
Identities = 35/105 (33%), Positives = 55/105 (52%)

Query: 145 SEVSQDHETNAQVLPFHHMLVALTFWKKRISIELVAYLLRIEDLQVVDCLPLWTL 204
S++ + RTHM VL SK+ G+ + I P + I DS VVVD S + W +
Sbjct: 489 SQIRKCHOTNCVLTSEKHLUTVRAVWVNGDKITSDSAVAIVHLSVVVLL----WIV 568

Query: 205 QEKQYISLGCYVOLLPLVKELLESFEETVIVLMLVLQAVIKRW 249
+ + L C -L+ -- LLSAE TV D L+ ---R+
Sbjct: 545 HQKASLWKLCTTPKQTERLGGSTETVQCTELKILQRF 589

Peptide information for DFFPhtes3_9k22, frame 3

Report for DFFPhtes3_9k22.3

[LENGTH] 304
[MW] 34767.24
[pI] 5.18
[SW] All_Alpha

WO 01/12659

PCT/IB90/01496

```
[SW]          LOW_COMPLEXITY      1.95 %

SEQ  HASETHVYKRMPCRIEDHFDLPRKLSMFTWHRKRYKSLPKLAATIRHTVQTVK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPDKLNVLYRRKEVHMFPHPCYTRQSPGSGGCHAKFKELACAGHLPKELIHDSRT
SEG  .....
PRD  GGNHNNNNNNNNNNcccccccccccccccccccccccccccccccccccccccc

SEQ  YLVHSIDGCSQTESPESKTBGFFFEYQDNEHQAQLFERNHMLVALTWKASISEL
SEG  .....
PRD  #####cccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VAYLLRISGKGVVQGLPVLTHCQGESQYISAGCCVLLPLVESLLEKFEETVIVGLA
SEG  .....
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  VLQAVIKRWSELSSTETIINQGHQIILKQGLGLWQDHLVLVPTQTHIARDVQATL
SEG  .....
PRD  hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  LQLH
SEG  ...
PRD  NCC
```

(No Prosite data available for DFFptex3_9422.3)
(No Pfam data available for DFFptex3_9422.3)

Localisation of expressed proteins

GeneID	Accession	Accession (top)	Accession (top)	Accession (top)	Accession (top)	Protein Group	Similarity	Chromosomal Location	Localization Predicted	Localization
U017466	U017466	640	5386	1773		transport and like protein	similar to kinesin like protein	912.5 kb from top of Chr10 linkage group	"nec"	nec
U017518	U017518	6955	1236	6035	1300	differentiation chromatin protein	similar to Drosophila chromatin protein		"no predict"	nec
U018289	U018289	2723	45	1380	413	signaling & communication	similar to myosin A-actin transferase	16	"Mitochondria"	Mitochondria
U017420	U017420	1376	132	612	167	unknown	unknown	4	"no predict"	Endoplasmic Reticulum
U018283	U018283	2314	76	1393	626	differentiation & development	similar to MD21 containing three zinc finger motifs present in OTF- binding protein, but not conserved in 2, 2x3.1	218.7 kb from top of Chr18 linkage group	"no predict"	nec
U018243	U018243	2892	317	1379	621	signaling & communication	similar to origin recognition complex	64.4, 3-14.1	"nec"	Cytosol + nucleus
U018243	U018243	2892	317	1379	621	signaling & communication	similar to protein involved in energy metabolism	10	"Mitochondria"	Endoplasmic Reticulum
U018243	U018243	2892	317	1379	621	signaling & communication	unknown, contains 2 WD-40 repeats, which are transmembrane subunit of C-protein		"no predict"	Cytosol + nucleus
U018243	U018243	2892	317	1379	621	signaling & communication	similar to RNA helicase		"nec"	nec

Clone ID	Accession	Gene (bp)	ORF (bp)	ORF (aa)	Protein group	Similarity	Translocation	Localization Predicted	Localization
DM7p548B162	ALL36466	659	641	213	535	dephosphorylation	4	"no predict"	Cytoplasm
DM7p548B122	ALL49972	1842	70	133	448	unknown		"no predict"	Cytoplasm
DM7p548B1671	ALL36467	1484	78	122	82	unknown		"no predict"	Endoplasmic reticulum
DM7p548B162	ALL36421	1914	246	463	463	nucleic acid metabolism	1912	"no predict"	other/unknown
DM7p548B163	ALL36468	1208	191	377	139	transport and traffic		"no predict"	Cytoplasm + Nucleus
DM7p548B212	ALL36423	1915	218	148	179	protein metabolism	23212.1	"no predict"	Endoplasmic reticulum
DM7p548B233	ALL36413	1300	121	699	193	signaling & communication	931.6 as from top of Chrt linkage	"no predict"	Cytoplasm + Nucleus
DM7p548B246	ALL36464	2084	73	187	134	metabolism	5	"no predict"	Nucleus
DM7p548C0462	ALL36476	1731	68	142	161	nucleic acid metabolism		"no predict"	other/unknown
DM7p548C0469	ALL36479	899	86	98	270	unknown		"no predict"	Cytoplasm + Nucleus
DM7p548C1362	ALL36447	437	137	473	179	metabolism	16312.3-213.11	"no predict"	other/unknown

Clone ID	Accession	Contig (bp)	Offset (bp)	Offset (bp)	Offset (bp)	Protein (aa)	Similarity	Chromosomal location	Localization Predicted	Localization
DFZ2564C1516	AL134597	3928	240	1987	586	Structure & mobility	shares the features of myosin and hunch and involved in the organization of cytoskeleton and to membrane proteins		Cytoplasmic / plasma membrane	Nucleus
DFZ2564C1512	AL134627	2305	155	425	157	membrane protein	unknown	86.2 kb from top of Chr1 linkage group	no predict*	endoplasmic reticulum
DFZ2564C1544	AL134654	1866	180	1540	287	unknown	unknown	79.8 kb / 79.8 kb / 812 kb	no predict*	Cytosol
DFZ2564C1512	AL134628	2835	272	1177	320	unknown	unknown		no predict*	Cytol
DFZ2564C1513	AL134639	1709	105	1448	448	nucleic acid management	similar to: DBD-box protein	87.58 kb from top of Chr1 linkage group	no predict*	Nucleus / nuclear envelope
DFZ2564C1516	AL134629	2256	166	166	280	signaling & communication	similar to: neuronal calcium sensor		no predict*	Nucleus
DFZ2564C1516	AL134622	2535	29	1449	627	signaling & communication	similar to: GTP-binding protein		no predict*	Cytosol
DFZ2564C1522	AL134631	1787	18	144	329	unknown	unknown		no predict*	Cytosol
DFZ2564C1523	AL134631	2005	104	1000	299	unknown	unknown	1413	no predict*	Cytosol + Nucleus
DFZ2564C1542	AL134657	2322	163	1581	473	signaling & communication	similar to: calmodulin-related protein	206.5 kb from top of Chr1 linkage group	no predict*	Cytosol + Nucleus

Clone ID	Accession	Length (bp)	Start (bp)	End (bp)	Protein group	Protein description	Similarity	Localization	Localization prediction
DF72p44E1712	AL134596	1619	40	732	311	Membrane protein	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	135	704	304	Cell Cycle	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	46	613	182	unknown	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	141	1323	123	signaling & communication	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	281	943	220	unknown	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	250	318	223	differentiation	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	115	718	288	unknown	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	156	1886	167	unknown	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	98	730	313	unknown	unknown	Top of Chromosome	"no predict"
DF72p44E1712	AL134596	1619	157	933	259	unknown	unknown	Top of Chromosome	"no predict"

CloneID	Access	Contig (bp)	ORFStart (bp)	ORFEnd (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation	Localization Predicted	Localization
DFP2646083	AL136641	10227	37	570	178	protein management	similar to: yeast, ARO1 and MCL1 are required for the import of an E-terminal protein acetyltransferase 1.	20	"no predict"	Cytosol + Nucleus
DFP2646082	AL136612	24444	559	1235	239	unknown	unknown	4922, 1-22	"no predict"	Cytosol + Nucleus
DFP2646083	AL136613	987	93	632	140	unknown	unknown		"no predict"	mitochondrion
DFP264608123	AL136699	1734	159	1133	325	membrane protein	unknown	11614	"no predict"	Nucleus
DFP264608232	AL136696	2292	270	1829	520	membrane protein	unknown	11612, 2 from Nucleus, 1182317	"no predict"	Cytosol + Nucleus
DFP264608152	AL136649	2014	75	971	299	structure & mobility	similar to: cell cell interaction protein	1	"no predict"	Plasma membrane + Nucleus + Cytoplasmic space
DFP264610323	AL136615	1467	126	1084	313	signaling & communication	similar to: protein activator of the nuclear envelope protein kinase		"no predict"	Cytosol
DFP264610422	AL136607	4748	511	1194	278	signaling & communication	unknown		"no predict"	Golgi + Plasma membrane
DFP264611316	AL136690	1548	81	635	185	membrane protein	unknown	877, 3-875, 1 of from top linkage group	"no predict"	Endoplasmic Reticulum
DFP264611782	AL136699	1761	348	610	81	signaling & communication	similar to: phospholipase protein.	11023	"no predict"	Golgi + Plasma

Clone ID	Accession	Gene ID (bp)	ORF start (bp)	ORF end (bp)	Protein group	Similarity	Chemical composition	Localization Predicted	Localization
DFPp441206	AL336465	1122	34	921	286	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441207	AL336466	1713	98	842	275	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441208	AL336467	1860	10	1650	547	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441209	AL336468	1409	9	1021	220	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441210	AL336469	2868	32	1639	496	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441211	AL336470	600	109	648	180	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441212	AL336471	1858	154	1460	429	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion
DFPp441213	AL336472	2775	719	2392	518	unknown	177.5 kD from top of Chr11 (large protein)		mitochondrion

Clone ID	Access	Contig (bp)	ORF start (bp)	ORF end (bp)	ORF size (aa)	Protein Group	Similarity	Chromosomal map	Localization Predicted	Localization
DFP254420822	AJ134419	2789	18	535	172	unknown	unknown	7	no predict*	membrane
DFP254423216	AJ134423	1978	397	1118	384	signaling & communication	similar to: GTP-binding regulatory protein	1	membrane*	Org. + membrane
DFP254423192	AJ134419	1971	187	1015	103	unknown	unknown	922.1-21.3	no predict*	Other/unknown
DFP254423244	AJ134419	1960	207	844	234	unknown	unknown	17	no predict*	Endoplasmic reticulum
DFP254423216	AJ134422	2048	232	1155	104	unknown	unknown		no predict*	mitochondria
DFP25442323	AJ134423	2978	279	2345	549	Protein management	Unknown, P/ra protein, ubiquitin	9	Cytosol*	Cytosol + nucleus
DFP254423216	AJ134423	2042	73	873	267	membrane protein	unknown		membrane protein*	Org. + membrane
DFP25442323	AJ134427	2414	29	1372	348	transport and cyclic	Unknown, a lectin character is predicted	3	membrane protein*	Endoplasmic reticulum
DFP25442323	AJ134427	952	227	949	121	unknown	Unknown, contains transmembrane helix		no predict*	Cytosol + membrane
DFP25442323	AJ134427	2406	14	315	104	signaling & communication	unknown	10.7-26.7 from N-terminus linkage group	no predict*	Org. + membrane
DFP25442323	AJ134424	614	34	429	125	unknown	similar to: Juncus protein		no predict*	Cytosol + nucleus
DFP25442323	AJ134422	1192	128	1027	301	signaling & communication	similar to: phospho- kinase protein, G-protein	9	Cytosol*	Cytosol

Clone ID	Accession	Gene (bp)	ORF (bp)	ORF (aa)	Protein (aa)	Protein (kDa)	Similarity	Localization	Localization
							mod1252		
DFP5640182	AL136217	299	323	110	116	116	unknown	no predict	Cytoplasm
DFP5640183	AL136218	2201	84	134	137	137	unknown	"cyto"	Cytoplasm
DFP5640184	AL136219	144	75	14	14	14	similar to D61 gene	"cyto"	Cytoplasm
DFP5640185	AL136220						unknown, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640186	AL136221	3336	172	187	192	192	similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640187	AL136222						similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640188	AL136223	2515	186	199	201	201	similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640189	AL136224	1736	24	110	110	110	similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640190	AL136225	1985	234	872	213	213	similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm
DFP5640191	AL136226	1260	56	941	332	332	similar to D61 gene, contains a M domain with binds to a particular protein.	"cyto"	Cytoplasm

[illegible]

GeneID	Accession	Species (sp)	Protein (aa)	GenBank (aa)	ProteinGroup	Similarity	ProteinID	ProteinName	ProteinFunction	ProteinLocation
AF036812	AF036812.1	Human	226	226	Transmembrane protein	Similar to: transmembrane protein	AF036812	Transmembrane protein	Transmembrane protein	Transmembrane protein
AF036813	AF036813.1	Human	186	186	Protein	Similar to: protein	AF036813	Protein	Protein	Protein
AF036814	AF036814.1	Human	186	186	Protein	Similar to: protein	AF036814	Protein	Protein	Protein
AF036815	AF036815.1	Human	186	186	Protein	Similar to: protein	AF036815	Protein	Protein	Protein
AF036816	AF036816.1	Human	186	186	Protein	Similar to: protein	AF036816	Protein	Protein	Protein
AF036817	AF036817.1	Human	186	186	Protein	Similar to: protein	AF036817	Protein	Protein	Protein
AF036818	AF036818.1	Human	186	186	Protein	Similar to: protein	AF036818	Protein	Protein	Protein
AF036819	AF036819.1	Human	186	186	Protein	Similar to: protein	AF036819	Protein	Protein	Protein
AF036820	AF036820.1	Human	186	186	Protein	Similar to: protein	AF036820	Protein	Protein	Protein
AF036821	AF036821.1	Human	186	186	Protein	Similar to: protein	AF036821	Protein	Protein	Protein
AF036822	AF036822.1	Human	186	186	Protein	Similar to: protein	AF036822	Protein	Protein	Protein
AF036823	AF036823.1	Human	186	186	Protein	Similar to: protein	AF036823	Protein	Protein	Protein
AF036824	AF036824.1	Human	186	186	Protein	Similar to: protein	AF036824	Protein	Protein	Protein
AF036825	AF036825.1	Human	186	186	Protein	Similar to: protein	AF036825	Protein	Protein	Protein
AF036826	AF036826.1	Human	186	186	Protein	Similar to: protein	AF036826	Protein	Protein	Protein
AF036827	AF036827.1	Human	186	186	Protein	Similar to: protein	AF036827	Protein	Protein	Protein
AF036828	AF036828.1	Human	186	186	Protein	Similar to: protein	AF036828	Protein	Protein	Protein
AF036829	AF036829.1	Human	186	186	Protein	Similar to: protein	AF036829	Protein	Protein	Protein
AF036830	AF036830.1	Human	186	186	Protein	Similar to: protein	AF036830	Protein	Protein	Protein
AF036831	AF036831.1	Human	186	186	Protein	Similar to: protein	AF036831	Protein	Protein	Protein
AF036832	AF036832.1	Human	186	186	Protein	Similar to: protein	AF036832	Protein	Protein	Protein
AF036833	AF036833.1	Human	186	186	Protein	Similar to: protein	AF036833	Protein	Protein	Protein
AF036834	AF036834.1	Human	186	186	Protein	Similar to: protein	AF036834	Protein	Protein	Protein
AF036835	AF036835.1	Human	186	186	Protein	Similar to: protein	AF036835	Protein	Protein	Protein
AF036836	AF036836.1	Human	186	186	Protein	Similar to: protein	AF036836	Protein	Protein	Protein
AF036837	AF036837.1	Human	186	186	Protein	Similar to: protein	AF036837	Protein	Protein	Protein
AF036838	AF036838.1	Human	186	186	Protein	Similar to: protein	AF036838	Protein	Protein	Protein
AF036839	AF036839.1	Human	186	186	Protein	Similar to: protein	AF036839	Protein	Protein	Protein
AF036840	AF036840.1	Human	186	186	Protein	Similar to: protein	AF036840	Protein	Protein	Protein
AF036841	AF036841.1	Human	186	186	Protein	Similar to: protein	AF036841	Protein	Protein	Protein
AF036842	AF036842.1	Human	186	186	Protein	Similar to: protein	AF036842	Protein	Protein	Protein
AF036843	AF036843.1	Human	186	186	Protein	Similar to: protein	AF036843	Protein	Protein	Protein
AF036844	AF036844.1	Human	186	186	Protein	Similar to: protein	AF036844	Protein	Protein	Protein
AF036845	AF036845.1	Human	186	186	Protein	Similar to: protein	AF036845	Protein	Protein	Protein
AF036846	AF036846.1	Human	186	186	Protein	Similar to: protein	AF036846	Protein	Protein	Protein
AF036847	AF036847.1	Human	186	186	Protein	Similar to: protein	AF036847	Protein	Protein	Protein
AF036848	AF036848.1	Human	186	186	Protein	Similar to: protein	AF036848	Protein	Protein	Protein
AF036849	AF036849.1	Human	186	186	Protein	Similar to: protein	AF036849	Protein	Protein	Protein
AF036850	AF036850.1	Human	186	186	Protein	Similar to: protein	AF036850	Protein	Protein	Protein

Clone ID	Accession	Gene ID	Protein ID	Protein Name	Protein Group	Similarity	Localization	Protein Name	Protein Group
DFPp1602019	AL02019	1782	266	1116	171	unknown	cytosol	cytosol	cytosol
DFPp1602018	AL02018	1076	65	516	184	similar to: mitochondrial protein	mitochondrial	mitochondrial	mitochondrial
DFPp1602017	AL02017	1886	23	1885	411	similar to: transport and	transport and	transport and	transport and
DFPp1602016	AL02016	1897	149	1701	454	similar to: transmembrane	transmembrane	transmembrane	transmembrane
DFPp1602015	AL02015	2275	79	1899	433	unknown	cytosol	cytosol	cytosol
DFPp1602014	AL02014	2638	84	1881	342	unknown	cytosol	cytosol	cytosol
DFPp1602013	AL02013	6892	187	1413	1159	similar to: GTP-binding	GTP-binding	GTP-binding	GTP-binding
DFPp1602012	AL02012	4117	187	1418	444	similar to: ADONIS	ADONIS	ADONIS	ADONIS
DFPp1602011	AL02011	4130	139	1894	1252	similar to: gpi	gpi	gpi	gpi
DFPp1602010	AL02010	1328	179	1445	442	similar to: gpi	gpi	gpi	gpi
DFPp1602009	AL02009	4293	112	1421	770	similar to: signaling &	signaling &	signaling &	signaling &

Table of cDNA clones and related data

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Group Nucleic acid management

[illegible]

[illegible]

[illegible]

Group signal transduction

[illegible]

[illegible]

1021

Group transcription factors

RelA	RelA is a transcription factor, composed of 315 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
RelB	RelB is a transcription factor, composed of 273 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
NF- κ B	NF- κ B is a transcription factor, composed of 500 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
RelA	RelA is a transcription factor, composed of 315 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
RelB	RelB is a transcription factor, composed of 273 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
NF- κ B	NF- κ B is a transcription factor, composed of 500 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
RelA	RelA is a transcription factor, composed of 315 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
RelB	RelB is a transcription factor, composed of 273 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor
NF- κ B	NF- κ B is a transcription factor, composed of 500 amino acids. It is a member of the NF- κ B family of transcription factors. It is involved in the regulation of the immune response and is a key component of the NF- κ B signaling pathway. It is a transcription factor that binds to DNA and activates transcription.	Transcription factor

1023

Profile Key

[illegible]

NAME: Cyclic nucleotide-binding domain signature 2.
CONSENSUS: [LIVMF]G-E-(GAH)[LVFM]-K(11)-R-[STAQ]-A+(LVNMA)+(STACV).
NAME: cAMP/GMP binding motif.
NAME: EF-hand nucleotide-binding domain.
CONSENSUS: D-(DHS)[LIVFYW]+(EENSTG)(DHQGHK)(CP)[LVFMQ](DEKQITAK)K2.
CONSENSUS: [DEH][LVFMFYW].
NAME: Actin-type actin-binding domain signature 1.
CONSENSUS: [EQ]K2H(LATV)QTY-K2-W-KM.
NAME: Actin-type actin-binding domain signature 2.
CONSENSUS: [LVFM]+(SDG)[LVFM](DAGREH)SAGS+(DNEAG[LVFM]+(DEAG)K6).
CONSENSUS: [LVFM]+(LM)SAGS[LVFM](LVMT)-W+(LVFM)S.
NAME: Amphiphysin domain signature.
CONSENSUS: [C]M[C-K]GAP-K7-R-[GATDSQR]-C-[GATDSQK]-K3-R-[GATDSQK]K2.
CONSENSUS: [CE]-K6-7-C-C.
NAME: Amphiphysin domain profile.
NAME: Apple domain.
CONSENSUS: C-K3-[LVMPY]K3-[LVMPY]K3-(DEMQ)[LVMPY]K3D-C-K3-C-T.
CONSENSUS: K4K-C-[LVMPY]F-K-[PT]K13K4C-[LVMPY]K3K3-[ST]K13-LD.
CONSENSUS: S-Q-[ST][LVMPY]K3D-C.
NAME: Reel 4.1 family domain signature 1.
CONSENSUS: W-[LVY]K7H(SL)S+[LVFM]K2H(QH)-K6-L3[LVMPY]K6-R[LVMPY].
CONSENSUS: K3-L3-FQTY-K3D-(DEDS).
NAME: Reel 4.1 family domain signature 2.
CONSENSUS: [H]VYK7H(SD)HQTYS(SA)-K3-PT[LVFM]K7H(ACV)-K3(LM)-K3.
CONSENSUS: [PT]G-S-(SD)QT[LVMPY].
NAME: Reel 4.1 family domain profile.
NAME: C16 domain signature.
CONSENSUS: F-K25-[QD]-K44[PTW]L2-K69-F-K23-D-K-T-K-F-K-[PT].
NAME: C-terminal cysteine loop signature.
CONSENSUS: C-C-K13K-C-K23(SH)K13D-C-K-K23-K1C.
NAME: C-terminal cysteine loop profile.
NAME: CUB domain profile.
NAME: Death domain profile.
NAME: EGF-like domain signature 1.
CONSENSUS: C-K-K3K-K3D-C.
NAME: EGF-like domain signature 2.
CONSENSUS: C-K-K2-[QY]PTW-K4E-C.
NAME: Calcium-binding EGF-like domain pattern signature.
CONSENSUS: [DSQY]+(DSQY)D2-C-K3-L4-C-K3-7-C-K-(DN)-K44-PT)+-C.
NAME: Laminin-type EGF-like (L2) domain signature.
CONSENSUS: C-K3-L2-K-K5Q-K23-C-K3-K3HPTW-K3-L3K-C.
NAME: Complement factors 5/9 type C domain (FAM3C) signature 1.
CONSENSUS: [GAD]W-K7-L3HPTW[LVY]+[LVFA]-HSTDEN-K66[LVFY]K23-PT)+-
CONSENSUS: [LVY]TSDKXG.
NAME: Complement factors 5/9 type C domain (FAM3C) signature 2.
CONSENSUS: F-K18-KH(LM)-K-(GE)[LVFY]+G-C.
NAME: Pankhoed-oxenoid (FHA) domain profile.
NAME: Fibronectin type and gamma chain C-terminal domain signature.
CONSENSUS: W-W-[LVMPY]K23-C-K23(GSA)-K23-N-G.
NAME: Type I fibronectin domain.

VO 01/13659

PCT/IB2001/13659

CONSEQUENCE: C=6k3,LFPT=C4=PTFY=C1=3k2,4,5,10=C3=C4=Ck3.

NAME: Type II cytochrome oxidase binding domain.

CONSEQUENCE: C=C2=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Hemoglobin domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Erythroid domain signature.

CONSEQUENCE: (PT)C=C8=N-F(DH81)

NAME: Erythroid domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: L2L-complex class A (L2LAA) domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: L2L-complex class A (L2LAA) domain profile.

NAME: C-type lectin domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: C-type lectin domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Last domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Osmotically sensitive domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Osmotically sensitive domain signature 2.

CONSEQUENCE: P=C8=N-F(DH81)

NAME: Symplectin-1 domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Symplectin-1 repeat signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: P-type lectin domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Carboxin-binding domain, internal type.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Carboxin-binding domain, external type.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Citrus recognition or binding domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Barren domain signature 1.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Barren domain signature 2.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: BIR repeat.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: WAP repeat.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: Pseudo-symplectin/haemoglobin binding domain.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

NAME: C1 domain signature.

CONSEQUENCE: C=C1=PTFY=C7=C3=C4=1b,16=C4=C41(DPFW3)C1,3,13,13,PTFY=C1.

PCT/IB00/01496

1027

NAME: Poly(ADP-ribose) polymerase zinc finger domain signature.
CONSENSUS: C-(KQ)+C- α 23+ α E- α 23-(KQ-N18.18) W-(PTVH)-H- α 23-C.
NAME: Poly(ADP-ribose) polymerase zinc finger domain profile.
NAME: Pungal Zn(2)-Cys(8) binuclear cluster domain signature.
CONSENSUS: (D)STPVK-C- α 23-C-(KQSTTACW)- α 23-(KRAHQ)- α 23-C- α 3.123-C- α 23-C- α 8.19.
CONSENSUS: C.
NAME: Pungal Zn(2)-Cys(8) binuclear cluster domain profile.
NAME: Peckysine dRNAse3 C4-cys zinc finger.
CONSENSUS: C-(EES)+C- α 23+ α 23-E- α 15-P- α 19-C- α 23-C.
NAME: Copper for domain signature.
CONSENSUS: M-(LVMQ)7D- α 23-E-(MY)-A-C- α 23-C-1-(K3)+H-(K3)- α 23-C-1-H- α 23.
CONSENSUS: (L2)+-(L2)-G-R-F.
NAME: Copper for DNA binding domain profile.
NAME: Lactase copper pattern.
CONSENSUS: L- α 16-L- α 18-L- α 23-L.
NAME: hZF transcription factor zinc domain signature.
CONSENSUS: (K3)+ α 1.23-(K3AAQ)-H- α 23-(SAQK2)+-(RTADMQ)+R+-(K3).
NAME: Myb DNA binding domain repeat signature 1.
CONSENSUS: W-(IT)- α 23-E-(DE)- α 23-(LV).
NAME: Myb DNA binding domain repeat signature 2.
CONSENSUS: W- α 23-(L2)-(AAQ)- α 1.23-E- α 23-(TW)- α 23-(LVFM).
NAME: Myo- α 23, "beta-loop-beta" disordered domain signature.
CONSENSUS: (GKNTA)P-L-(LVFM)A(ED)-1-(PTW)CH-R-(LVY)H(LV)- α 23-(STA)Y-(LVMASTAC)+
CONSENSUS: (VMTY)S-(LVMTA)-1-(Y)-1-(LVMAE).
NAME: p53 tumor antigen signature.
CONSENSUS: M-C-R-S-S-C-M-Q-G-M-P-R-E.
NAME: CBP-A/P2-VB subunit signature.
CONSENSUS: C-V-S-S- α 15-F-(LVFM)-Y-(Q)-E-A-(K)-DM-(K)Q-C.
NAME: CBP-A/P2-VB subunit signature.
CONSENSUS: Y-V-M-A-E-Q-Y- α 1-L-E-S-S- α 15-A-R-A-E-L-E.
NAME: "Cold shock" DNA binding domain signature.
CONSENSUS: (PT)-G-P- α 16.7-(DE3)-(LVMA)P- α 15+-(PTK3)+-(LVMAFY).
NAME: CTF/TF1 signature.
CONSENSUS: E-E-R-E-Y-P-E-E-R-S-E.
NAME: Ets-domain signature 1.
CONSENSUS: L-(PTW)Q(EDM)-P-D-E-(LVQ3)+-(R)-L.
NAME: Ets-domain signature 2.
CONSENSUS: (R)S- α 23-M- α Y-(EDM)Q-(LVFM)-1-(TAD)- α 23-(TAGH)-R- α Y.
NAME: Ets-domain profile.
NAME: Fork head domain signature 1.
CONSENSUS: (K3)-P-(PTQ)-PTLVQ(R)-S-(P)- α 23-(LVFM)- α 23-A-(AC)-L(M).
NAME: Fork head domain signature 2.
CONSENSUS: W-(R2)-P(K3)- α 23-(LVF)R-H.
NAME: Fork head domain profile.
NAME: RSP- α 23 DNA binding domain signature.
CONSENSUS: L- α 23-(PT)S-E- α 15+-(ETAD)-S-P-(LVMA)S-Q-L-(H)- α Y+-(PTW)R(K2)S-E.
CONSENSUS: (LVFM).
NAME: Tyrosine protein repeat (RFP family) signature.
CONSENSUS: W- α 23(HS)- α 23-(LVFY)- α 23-P-H- α 15- α 19.16-(DE)- α 23-(LVFY)-P-(K)Q- α 15.

1029

PCT/IB2010/01496

1030

CONSENSUS: LLYMAA(LQRN)KZ(LVYMPW)KZ(LVYM).

NAME: *Sign-70 known ECF substrate signature.*
CONSENSUS: STTAIVLPKQSLIDRL(LVY)LVYTAI-Q-+STAVY(LVYMPYC)LVYKAAQ-
CONSENSUS: GGTATPY(LDAPFTWQ)K(LI)HSTATPY(PYVY)LPYKZ(DY).

NAME: *Sign-34 unknown domain ATP binding region A signature.*
CONSENSUS: LLYMPYED+G(DKQ)STTS-Q(STAVY)GK-KZ(LVMPY).

NAME: *Sign-34 unknown domain ATP binding region B signature.*
CONSENSUS: KGB+GLVMPYKZ+A(DNEDQASH)GHEK)G(PTM)(LVMPYK)DRE)EK).
CONSENSUS: LLYM).

NAME: *Sign-34 unknown domain C-terminal part signature.*
CONSENSUS: (PYVY)P(KG)M(LVYMG-K(SQ)L+PQHAT).

NAME: *Sign-34 unknown domain profile.*

NAME: *Single-strand binding protein family signature 1.*
CONSENSUS: LLYMPY(LNTT)KRT(LVYM)+GLVMPYD-G(NHAK)(LVYM)GRT+GSET).

NAME: *Single-strand binding protein family signature 2.*
CONSENSUS: T+M(DT)HMQ(LVYAC)-(LVYMPY)PT)HGX).

NAME: *Structural homology DNA-binding protein signature.*
CONSENSUS: KGBK)P-KZ(LVMPY)K+G(BKQDA)KZ(HSTY)+(GA)+(LYN)P+T.

NAME: *Dye protein family signature 1.*
CONSENSUS: M(PYVY)+(LVYM)+G-KZ(LVY)H-KZ(DK).

NAME: *Dye protein family signature 2.*
CONSENSUS: LLYMPY(DM)+(LVYM)(GA)S-R-KZ(LUP)KDN)KZ(PA).

NAME: *DNA repair protein recC family signature.*
CONSENSUS: M-P-K-P-S-Q.

NAME: *recA signature.*
CONSENSUS: A-L(EK)DPY(PY)STAI-STADN(LVYMQ)R.

NAME: *BacF protein signature 1.*
CONSENSUS: P(EK)KZ(LVYKED)+G(GSAD)P-KZ)R-R-KZ(PY)LVYMQ-D.

NAME: *BacF protein signature 2.*
CONSENSUS: LLYMPYKZ)+S-KZ)B(SA)HED)L-D-KZ(GMA)KZ)N-L.

NAME: *BacB protein signature.*
CONSENSUS: C-KZ-C-KZ(STY)K+K+C-KZ-KZ)R.

NAME: *HemK H2A signature.*
CONSENSUS: IAC)G-L-P-P-P-Y.

NAME: *HemK H2B signature.*
CONSENSUS: (EY)P+GLVY(LKQ-T-KZ)EK)+GLVYKZ)+(PAQ)DQ-L+(EK)H+A-
CONSENSUS: LLYMQ(STA)E-Q.

NAME: *HemK H3 signature 1.*
CONSENSUS: E-A-P-E-S-Q-L.

NAME: *HemK H3 signature 2.*
CONSENSUS: P-P+(SA)-Y(A)K(PQ-DKQ)PY).

NAME: *HemK H4 signature.*
CONSENSUS: G-A-E-E-E.

NAME: *HMG1/2 signature.*
CONSENSUS: (PY)S(EK)E-C-S(EK)S-W-E-T-M.

NAME: *HMG14 and HMG17 DNA-binding domain (A+T-rich).*
CONSENSUS: (AT)K(L)H(EK)D(GP)S-G-R-P(EK)+.

NAME: *HMG14 and HMG17 signature.*
CONSENSUS: R-S-S-R-S-A(EK)P.

NAME: *Structural signature.*

PCT/JP00/01496

1032

1033

PCT/IB02/01496

1034

NAME: Ribosomal protein S11 segment.
CONSENSUS: (LIVMF)₅+GTTACT(LIVMF)₅+C2(GGTAL)₄+W₁(1)GSD(LIVMF)₅+LIVM;
CONSENSUS: K(C)(DEW)₅+T-P+Q(A)(STCN)₃(DH).

NAME: Ribosomal protein S12 segment.
CONSENSUS: (E2)₂+P-R-L(A)₃+R.

NAME: Ribosomal protein S13 segment.
CONSENSUS: (E2)(S)₂-Q+R-R- α (2)-GSDW₅+C2(LIVMC)₂R-G-Q.

NAME: Ribosomal protein S14 segment.
CONSENSUS: (K7)₅+W₁-C- α (1-12)(LIVMF)₅+LIVMF₅(S2)₂(RQ)₂+D₁(P).

NAME: Ribosomal protein S15 segment.
CONSENSUS: (LIVM)₄+C2W-LIVMF₅+ α (3)-D- α (2)(SAQ)₂+C3(L7)₂+W₁(LIVM)₄+C3;
CONSENSUS: (PT).

NAME: Ribosomal protein S16 segment.
CONSENSUS: (LIVMT)₃+LIVM(LK)₃-L-STTAS₂R- α -G-(AX)₂.

NAME: Ribosomal protein S17 segment.
CONSENSUS: G-D+L(V)₅+L(V)₄+ α (2)E2+ α (2)E2-P(LV)₃R.

NAME: Ribosomal protein S18 segment.
CONSENSUS: (V)₂+D(T)₂-Y- α (2)-L(V)₂MT₂+C3(LIVM)₄+C3(GPTT)₂(LIVM)₂(ST)₂(DEPT)₂+
CONSENSUS: (GTY)₂(LIVM)₄+D₁R-LIVMAA₃.

NAME: Ribosomal protein S19 segment.
CONSENSUS: (STTMS)₂-G-(S)₂(DH)₂+W(LIVM)₄+W(LIVM)₂(GSD)₂+C7(L7)₂(SA)₂(DH)₂P;
CONSENSUS: α (2)-D(T).

NAME: Ribosomal protein S21 segment.
CONSENSUS: (DH)₂+ α -L(T)₂(K)₂R-P-E-(K)₂+C3(K).

NAME: Ribosomal protein S24 segment.
CONSENSUS: (L7)₂+GSD₂(-D)₂+E+ α (2)₂L- α -D-L.

NAME: Ribosomal protein S26 segment.
CONSENSUS: R₁+E+L(VM)₂(SA)₂+P- α (2)W- α (LIVM)₂+ α (K).

NAME: Ribosomal protein S28 segment.
CONSENSUS: (LIVM)₂(STTMA)₂+G-Q+D- α (2)-G+P-M.

NAME: Ribosomal protein S7a segment.
CONSENSUS: (K)₂L+R-S-L-E-E-P(LA)₂+ α (K)₂H.

NAME: Ribosomal protein S7b segment.
CONSENSUS: R- α (2)-T-G-(D)₂+ α (2)(P)₂R-(K)₂+R+R+R(LM)₂G.

NAME: Ribosomal protein S7c segment.
CONSENSUS: A-L(LS)₂(P)₂+ α -L- α (2)(SA)₂+C3(DH)₂-G-L.

NAME: Ribosomal protein S7d segment.
CONSENSUS: A+ α (ST)₂E+L-R-R-(K)₂+ α -G-(PT)₂+T-H.

NAME: Ribosomal protein S7e segment.
CONSENSUS: P- α (LSA)₂+C3(LIVMA)₂+R+ α (LIV)₂(LV)₂-Q+L(D).

NAME: Ribosomal protein S21a segment.
CONSENSUS: L-T-V-P-R-E-C-(SA).

NAME: Ribosomal protein S24a segment.
CONSENSUS: (FA)₂-Q- α (2)(LSTAA)₂+G-(PT)₂(SA)₂+L(VM)₂-Y-DH₂(SD).

NAME: Ribosomal protein S26a segment.
CONSENSUS: (TH)₂-C-V-S-C-A-H.

NAME: Ribosomal protein S27a segment.
CONSENSUS: (K)₂-C- α (2)-C- α (2)-P-(SA)₂+ α (2)-C-(K)₂+ α (2)-L- α (2)-P+G.

NAME: Ribosomal protein S28a segment.
CONSENSUS: R-(ST)₂S-S-R-S- α -L.

NAME: DNA minicircle repeat protein mod. / mod. / PHE1 segment.

WO 01/12659 PCT/JP00/01496

[illegible]

PCT/IB08/01496

1037

WO 81/13659

PCT/BB9001496

NAME: Aldehyde dehydrogenase: glucose acid kinase iso.
CONSENSUS: [LTVWDAALSLNETAC]KSDG-LSKLM] (SADN) (TAPPY)
NAME: Aldehyde dehydrogenase: cytosolic iso.
CONSENSUS: [PTLVAY]KTS-G-IGQ3 + C-[LVYMGTAHCT]AQCCH + RSTADHEKX]
NAME: Aspartate aminotransferase dehydrogenase isozyme.
CONSENSUS: [LVYH]EADH]KZC + S-[LVYH]KQ-(GK)H (STTA)
NAME: Glutathione S-transferase dehydrogenase isoform.
CONSENSUS: [APY]S-C-(PT)T-KZD-ELM]
NAME: N-acetyl-glucosyl-phosphatase reduction isoform.
CONSENSUS: [LVYH]GSA3 + P-G-C-(PTT)SAVYI-T (DA)-K3H-TAC) (LVYH) + P.
NAME: G-protein coupled receptor isoform.
CONSENSUS: V-KD + GLP3 + H3 + CD (VY) (SD) (PT) + H (ST) (DQ) + L
NAME: Dihydrodipicolinate reductase isoform.
CONSENSUS: S-DV + S + M-KZ + E + D + P-S-G-T-A.
NAME: Dihydrodipicolinate reductase isoform 1.
CONSENSUS: [GDS + LVA]GSA3 (TTA) (PTTA) (GT) + KZ (HQH) + G (HMH) + D + P (RT)
NAME: Dihydrodipicolinate reductase isoform 2.
CONSENSUS: [LVY]ST-(GSA3) + G-G-SVY3 + (PTQH) + KZ-(ACV) + H3-G-A.
NAME: Cytoplasmic triose isomerase isoform.
CONSENSUS: S + W-C-KZ + (PTK3) + (LVYH) + H + S + S-G-(LVYH) + G-G-(LVYH) + P-P-D.
NAME: Phosphate substrate / substrate dehydrogenase FAD-binding iso.
CONSENSUS: S-(PTT)H-(PTT)KZD + G-G
NAME: Acyl-CoA dehydrogenase isoform 1.
CONSENSUS: [GAC]S-LVYH-(PTT)-S-KZ-(GSAH) + G-PTT-D-KZ-(GSA)
NAME: Acyl-CoA dehydrogenase isoform 2.
CONSENSUS: [DQK]KZD-G-KDS-KZD-LVYH (PTT) + KZ-(DQK) + G-(K3E) + KZ-(DQK)
NAME: Aldehyde dehydrogenase & pyridine nucleotide transhydrogenase isoform 1.
CONSENSUS: G-(LVYH) + P + S-KZ + H-E-KZ + B-V-A + (PTT) + P + (GST) + Y-KZ + L + (K3KQ) + G
NAME: Aldehyde dehydrogenase & pyridine nucleotide transhydrogenase isoform 2.
CONSENSUS: [LVYH]KZD-G-KSA1-G + A-G-KZ-(SAB) + KZ-(GSA3) + (DQK) (LVYH) + G-A + V-KZ + G
NAME: Ole / Lin / Lin / Val dehydrogenase isoform iso.
CONSENSUS: [LVY]KZD-G-G-(SAG)K + KQVY-KZ-(DQST) (PL)
NAME: Oxidative acid substrate isoform.
CONSENSUS: [LVYH]KZD-H (DQK) + T-G + (GSA)KZ + G-KZ + G + A.
NAME: Pyruvate kinase isoform.
CONSENSUS: [LVYH]S-L-P-W (DQK) + KZ + S-LVYH-KZ-(DQK) + L
NAME: Copper metal substrate isozyme isoform.
CONSENSUS: [LVYH] (LVYH) + (LVYH) + G-T-KZ-H-Y-(DQK) (TY)
NAME: Copper metal substrate isozyme-binding isoform.
CONSENSUS: T + G-KZ-H-LVYH (PTT) + KZ + S-LVYH-KZ-(DQK) + P
NAME: Lipid oxidase substrate isozyme-binding isoform.
CONSENSUS: W-KZ-W-S-KZ-H-Q-H-Y-H
NAME: Delta 1-pyridine-3-carboxylate reductase isoform.
CONSENSUS: [PLPY]KZD-3-(LVYH) + KZ-(LVYH) (PTT) + (KAP) + G + T-KZ-(KAC)
CONSENSUS: [LVY]KZD-ELMP] (DQK)
NAME: Dihydrodipicolinate reductase isoform.
CONSENSUS: [LVYH]KZD-G-KZ + (LVYH) + P-W-KZ-(DQK) + KZ-(PTTY) + KZ-(PTTY)
NAME: Tetrahydrodipicolinate dehydrogenase / dihydrodipicolinate isoform 1.
CONSENSUS: [DQK] + (DQK) (LVYH)KZD-KZ-(LVYH) + KZ-(LVYH) + H + (DQK) + KZ-(LVYH) + H

PCT/TB00/01496

1039

1040

WO 81/12459

PCT/80061196

NAME: Manganese and iron superoxide dismutases signature.
CONSENSUS: D-W-S-H-E-T-A-I-L-P-T-T-T.

NAME: B-Branched-chain reduction large subunit signature.
CONSENSUS: W-W-D-E-L-F-W-K-L-T-G-E-L-V-M-E-F-T-A-I-Q-N-H-W-D-E-T-A-Q-L-V-M-H-A-S-C-K-W-D-
CONSENSUS: [P-A].

NAME: B-Branched-chain reduction small subunit signature.
CONSENSUS: [F-V-M-S-E-Q-S-W-L-D-E-L-Y-T-A]-[D-I-Y-E-S-I-A]-[E-T-A-Y-M]-Y-W-D-E-L-V-M-Q-W-D-
CONSENSUS: [L-P-T]-[D-F-P-C-A].

NAME: Hemoglobin component 1 alpha and beta subunits signature 1.
CONSENSUS: [L-V-M-F-T-S]-[L-V-M-F-T-T]-H-A-Q-[A-G-S-P]-[L-V-M-H-Q-A]-[A-Q-C].

NAME: Hemoglobin component 1 alpha and beta subunits signature 2.
CONSENSUS: [E-T-A-N-Q]-[E-T]-C-W-D-G-D-[D-H]-[L-V-M-T]-E-[E-T-A-C-A]-[L-V-M-F-T-T].

NAME: Hsp70/wc family signature 1.
CONSENSUS: R-W-C-C-F-W-D-S-A-A-W-G-C-[A-Q-G].

NAME: Hsp70/wc family signature 2.
CONSENSUS: D-L-L-G-D-V-V-C-G-G-P-[A-Q-S-P].

NAME: Nickel-dependent hydrogenase large subunit signature 1.
CONSENSUS: R-C-G-L-P-M-T-S-E-W-D-E-S-E-M-S-R-C-C-E-L-V-M-H-C.

NAME: Nickel-dependent hydrogenase large subunit signature 2.
CONSENSUS: [F-T]-D-P-C-E-B-H-[A-S-E-C-W-D]-N.

NAME: Chaperonin-60 kDa subunit signature.
CONSENSUS: H-E-L-V-M-W-D-E-L-P-M-E-[E-T-A-C-T]-[L-V-M-H]-[D-H-Q]-S-E-L-P-M-A-[L-V-M-Q-T]-[E-P]-S-
CONSENSUS: -[Q-W]-[Y-T]-[E-T]-[E-T-A-C]-[Q-L-V-M-H]-[E-R].

NAME: Bacterial-type phytylase dihydrogenase signature.
CONSENSUS: [H-Q]-[E-T-W-T]-[L-V-M-T]-[G-A-Q-C]-[E-S]-[T-A]-[H-M-T]-P-D-[E-T-A-Y]-G-E-L-V-M-
CONSENSUS: -W-D-E-S-L.

NAME: Glycine radical signature.
CONSENSUS: [E-T-V]-S-E-[Y-T-H-C-A]-G-T-[E-S-A-C-Y].

NAME: Brg1/actin/actin ERM/ERM2 family signature 1.
CONSENSUS: D-W-D-E-L-P-M-Q-T-G-W-E-T-T-S-W-K-D-A-P-P-R-L.

NAME: Brg1/actin/actin ERM/ERM2 family signature 2.
CONSENSUS: [L-V-M-Q-D-H-E-W-D-S-W-C-W-D-S-C-W-D-E-Y-O].

NAME: NRM/TPP/TPP/TPP family of acetyltransferase signature.
CONSENSUS: E-A-D-A-G-D-S-P-T-E-T-T-Y-Q-A-L-L-S-H-C.

NAME: RNA methyltransferase track family signature 1.
CONSENSUS: [D-N]-P-[P-A]-S-W-G-W-L-H-[L-V-M-E-D]-Y-S-E-C-H-W-D-T.

NAME: RNA methyltransferase track family signature 2.
CONSENSUS: [L-V-M-Q]-D-S-P-P-Q-T-T-[E-T]-W-E-L-[L-V-M-F-T]-E.

NAME: Thymidine kinase active site.
CONSENSUS: S-W-D-E-L-V-M-Q-W-T-[F-W]-[Q-W]-W-S-E-L-V-T-L-C-[E-L-A-Y-M]-W-T-[Q-M-T]-[F-Y-W]-
CONSENSUS: -E-L-V-S.

NAME: Ribosomal RNA subunit dihydrogenase signature.
CONSENSUS: [L-V-M-Q]-[L-V-M-F-T]-[E-T-A-C]-[G-A]-[L-V-M-F-T]-[E-T]-W-E-L-V-M-Q-
CONSENSUS: -W-E-L-V-M-T-[E-T-A-C]-[L-V-M-F-T]-[E-T-A-C]-[L-V-M-F-T]-[E-T-A-C]-[L-V-M-Q]-S-D.

NAME: Methylated DNA - cytosine-cytosine methyltransferase active site.
CONSENSUS: [L-V-M-F-T]-C-W-S-E-L-V-M-F-T-S.

NAME: H-4 Adenosine-specific DNA methylase signature.
CONSENSUS: [L-V-M-A-C]-[L-V-F-T-W-A]-[D-M]-P-P-[F-T-W].

NAME: H-4 cytosine-specific DNA methylase signature.
CONSENSUS: [L-V-M-F-T]-S-P-P-[F-T].

NAME: C-5 cytosine-specific DNA methylase active site.
CONSENSUS: [D-E-N-E-S]-[P-L-V]-W-T-[E-T-C]-P-C-W-T-[P-V-W-L-D-Q-S].

1042

WO 01/12659

PCT/TB00/01496

CONSENSUS: R-(LIVMFTS)-s-(LIVM)-s-(QMG)-s-G-C-(FYNA)-(GA)-C-(GA)-(ITAV)-s-(LIVMFTS)-
 CONSENSUS: (RA).

NAME: Myristoyl-CoA protein N-myristoyltransferase signature 1.
 CONSENSUS: E-L-N-F-L-C-L-H-K.

NAME: Myrtaoyl-CoA protein N-myristoyltransferase signature 2
 CONSENSUS: K-F-Q-L-G-D-Q

NAME: *Gemmus glaucostriatus* sp. nov.
 CONSENSUS: T-(STA)-H-(ST)-(LIYMA)-A(G)-(IN)-I-V-(STA)-I-T-I-T-(LIYM)-(NE)
 CONSENSUS: A(1,2)-(FY)-G.

NAME: Transglutaminase active site.
 CONSENSUS: **IoTI-Q-HCAI-W-Y-x-(3A-HGA)-(TVT)-x(2)-T-x-(LMSC)-R-(CSA)-LV-Q.**

NAME: Phosphorylase pyridoxal-phosphate transaminase
 CONSENSUS: E-A-I-K-Q-x-K-G-S-x-M-E-x-T-D-I-M-I-N.

NAME: UDP-glycoyltransferase signature
 CONSENSUS: [FW]-x(2)-Q-x(2)-[LYMYTA]-[LDWV]-x(4,8)-[LVGAC]-[VFPYA]-[LYMFP]-[TAGCM]-
 CONSENSUS: [HNQ]-[STAGC]-Q-x(2)-[TAGC]-x(3)-[STAGL]-[LYMFA]-x(4)-[PQR]-[LYMFT]-
 CONSENSUS: [D]-[PA]-x(3)-[DEG]-Q-[STW].

NAME: Purine/pyrimidine phosphoribosyl transferase signature
 CONSENSUS: (L)YMFYVCTA)-(L)YNG)-(L)YVMA)-(L)YVPC)-(DE)-D-(L)YMS)-(L)YVH)-(T)AYD-
 CONSENSUS: (T)AS)-(GAC)-x-(T)AS).

NAME: *Chlamydomonas reinhardtii* flag-tagged *atb*.
 COMMENTS: [PAB-LYMPY]-[LYMPY]-G-[LYMPY]-C-[LYMPY]-G-1-(ODS-1-[LYMPY]

NAME: Glyoxalate glutathione transferase class-B active site.
Coordinates: 6-48, 113-120, 170-171, 204-205, 240-241

NAME: Purine and other phosphorylation family 1 signature

NAME: Porcine and other phagocytin family 2 signature.

NAME: Thymidine and pyrimidine nucleoside phosphorylation inhibitors.

NAME: ATP phosphotransferase regulatory

NAME: NAD:arginine ADP-ribosyltransferase regulator

NAME: Polynesian blackboard sturgeon, *Acipenser melanostictus*

NAME: *1,4-bis(methylthio)butane dication* 1

CONCLUSION: G-A-G-D-G-C-A-G-Y.

CONSENSUS: O-(GA)-O-(ASC)-P-S-K-(DE).

CONSENSUS: (L-V)(2)-1-D-D-1(2,4)-D-R(4)-R-R-(2H)

CONSENSUS: [LVNMFY]-Q-KI-[FTL]-Q-(LVN)-I-Q-D-(LVNMFY)-I-(DNCI)

CONSENSUS: Y-(CSAM)-x(2)-(Y30)-A-(GSA)-(LIVAT)-(TV)-G-x(2)-(LMBC)-x(7)-(LIV)

NAME: Squamos and physocera synthetase signature 2.
 CONSENSUS: [LIVM]-G-a(3)-Q-a(2,3)-M-[IF]-s-R-D-[LIVMFPY]-a(2)-[DE]-a(4,7)-R-s-[F

NAME: ~~James Earl Ray~~ ~~aka alias name signature~~
$$f(x) = f_0 + f_1 x + f_2 x^2 + \dots + f_{n-1} x^{n-1} + f_n x^n - m(x^n - 1)$$

1043

NAME: *Radiololium syntrophus* alpha class family signature
CONSENSUS: [LVMF]c7y-G[STADMQ]E[ESQYV]Y V H [LVM]B

NAME: *Dibutyrate syntrophus* signature 1
CONSENSUS: [LVM]y + [AQG][LVMF]23 H + 7 + D S P + D + [E]

NAME: *Dibutyrate syntrophus* signature 2
CONSENSUS: [QD]SAy + [LVMKE]D[LVMQ]G[QF]wT[ST]A + P

NAME: *EFIP syntrophus* signature 1
CONSENSUS: [LVM]c7y[QW]R [SA]G-T[ST]A + R + [LVMY] + [KST]A

NAME: *EFIP syntrophus* signature 2
CONSENSUS: [TSA] + [QD]S-L[STT] [QK]G + [LVM] + [ST]A-[LVMK]c7y[DN]LVMFy +
CONSENSUS: [KLA]LVMFyG

NAME: *FLAPKISTLTCG* family signature
CONSENSUS: Q-c7y-D-S-V[PT]y + A-DQ3 + R-C

NAME: *Aminotransferase class-I pyridoxal-phosphate attachment site*
CONSENSUS: [QD][LVMFPTAC]K[ST]A-E-c7y[ESALVW]LVMFA) + [QNA]R + R [LVMMA]
CONSENSUS: [GA]

NAME: *Aminotransferase class-II pyridoxal-phosphate attachment site*
CONSENSUS: T[LVMFPT]STAG E-[SAQ][LVMFPTW]E[SAQ]c7y-[SAQ]

NAME: *Aminotransferase class-III pyridoxal-phosphate attachment site*
CONSENSUS: [LVMFPTWC]D + D-E[LVMMA]c7y[QF]wB-1-[LVMFPTW]ACD-wB-1-[SAQ]2 +
CONSENSUS: [ESAL]B-c7y-10-10-[LVMFPTWC]c7y-3-[SA]A-E-c7y-[STADN]E[SA]

NAME: *Aminotransferase class-IV signature*
CONSENSUS: E + [TAC]G-c7y-H-[LVMFA]C[PT]y-c7y-12-[LVMF]y + T-wB-D-[LVM] +
CONSENSUS: [QD]LVM + [ES]

NAME: *Aminotransferase class-V pyridoxal-phosphate attachment site*
CONSENSUS: [LVMFCTH]E[QW]E[LVMFPTAC]LVMFPTA) c7y[ESTAC]K[ST]A) [QW]E E-
CONSENSUS: c7y[B-G] + [SAAT] + [LVMFPTAC]

NAME: *Phosphotransferase signature*
CONSENSUS: [LVM]G-P [TN]R-S-PTT P-c7y-[LVM] [QW]T) c7y-[LVMQ]c7y-W-T-E-
CONSENSUS: [L]

NAME: *Catalase signature*
CONSENSUS: G-S + R [LVM]S-G-S-R + D-Y

NAME: *GDP 3'ase protein ATP binding domain*
CONSENSUS: [LVM]c7yK + [ST]A) wB-1)G-L[QD]S-S-[SA] [STAC]

NAME: *Phosphotransferase signature*
CONSENSUS: [SA]c7y-wB-1)G-[QW]G-G-c7y-D-R

NAME: *phd family of carboxylate kinase signature 1*
CONSENSUS: [AQG]G-wB-1)[SA]Fy + R + [ST]A) wB-1)[QD] wB-1)G

NAME: *phd family of carboxylate kinase signature 2*
CONSENSUS: [QW]L[PTT]y + [SAQ]D-[QD]D-c7y-[SAQ]Y[AG][LVMF]LVMETAF

NAME: *ROE family signature*
CONSENSUS: [LVM]c7yG-[LVMFCT]G + [SA]L[LVMFA] wB-G-c7y-[SAAT]c7y-
CONSENSUS: Q-D[ES]

NAME: *Phosphotransferase signature*
CONSENSUS: E [LVM]wB-S-D-c7y-G + [ST]y + E

NAME: *Threonine kinase cellular type signature*
CONSENSUS: [SA]c7y-[QD] + T + [ST]AF + C-[VSR] + [CTD]LVMFPTW

NAME: *FOOT family of carboxylate kinase signature 1*
CONSENSUS: [SA]FCTD + [ST]y-c7y-E [LVMFPT]y + W [LVMF]y + [QW]T[QD] [QW]Q

NAME: *FOOT family of carboxylate kinase signature 2*
CONSENSUS: [SA] + [LVMFPT]y + G [LVM]c7y-B-[QW]Q[LVMF]y-c7y-[SA] [ST]AFVW
CONSENSUS: [LVMFPT] [QW]Q

NAME: Protein kinase ATP-binding region signature.
CONSENSUS: [LPT]G-[PT]G-[PT]PTFMGQTNQ[RGSA][PM][LVCAT]-[PD]-[GDTACLVNMPY]
-K[ILH]L[LVNPTCT]A[LA]A[PTV]L[LVNMPACEKLS]K.
NAME: Serine/Threonine protein kinase active site signature.
CONSENSUS: [LVNMPYC]-[KTY]-D-[LVNMPY]-E-W[TH]M-[LVNMPYCTKLS].
NAME: Tyrosine protein kinase specific active site signature.
CONSENSUS: [LVNMPYC]-[KTY]-D-[LVNMPY]-[KSTAC]-K[TH]M-[LVNMPYCTKLS].
NAME: Protein kinase domain profile.
NAME: Casein kinase II regulatory subunit signature.
CONSENSUS: C-P-K[LVNMPY]-C-K[ILH]-P-[LVNAC]G-K[PS]V-[KSR]-K[TH]-C-P-K-C.
NAME: Pyruvate kinase active site signature.
CONSENSUS: [LVYAC]-[LVNMPY]-[KAPCV]K-[LVY]E-[HKEST]-[VDEQK]-[GSTA]L[LVNMPY].
NAME: Shikimate kinase signature.
CONSENSUS: [KLS]-K[TH]-E-K[ILH]L[LVNMPY]-K[ILH]L[LVNMPY]D-[KLS]-[GTH]-[LVNMPY].
NAME: Prokaryotic diacylglycerol kinase signature.
CONSENSUS: S-[LVNMPY]-N-[STT]SA-[LVY]-S-K[TH]-V-D.
NAME: Phosphatidylinositol 3- and 4-kinase signature 1.
CONSENSUS: [LVNMPAC]K-K-[ILH]D[SA]-[KSG]-[LVNAC]K-K-Q[DS]-K[TH]-Q.
NAME: Phosphatidylinositol 3- and 4-kinase signature 2.
CONSENSUS: [KSD]-[LA]Y-K[TH]-[LVNMPY]-K[TH]-[PTT]L[LVNMPY]D-[LVNMPY]-D-E-K-K-K[TH]-M.
NAME: Acetate and butyrate kinase family signature 1.
CONSENSUS: [LVNMPY]D-[LVNMPY]-N-K-G-S-[STT]-S-[KLS].
NAME: Acetate and butyrate kinase family signature 2.
CONSENSUS: [LVNMPY]D-K[TH]-K[ILH]-K[ILH]-G-[STT]-[LVNMPY]-[LA]Y-K[TH]-G.
NAME: Phosphoglycerate kinase signature.
CONSENSUS: [KLMKTCY]-[PTT]-[LVNMPY]-[LVNAC]K-K-[D]-N-[KACV]-P.
NAME: Aspartokinase signature.
CONSENSUS: [LVNMPY]-E-[STT]-G-D-[STT]-[KCT]-[LVNMPY].
NAME: Chemosome 3-kinase signature.
CONSENSUS: [KSTTQ]-K[TH]-G-K-[KCT]-[KMS]-[STT]A-E-[LVNMPY]-[SAH]-[TCA]-K[TH]-[KSA]YV.
CONSENSUS: -K[TH]-G.
NAME: ATP granule phosphotransferase active site.
CONSENSUS: C-P-K[ILH]-[STT]-K[ILH]-G-T.
NAME: PTD HPS conjugated histidine phosphorylation site signature.
CONSENSUS: G-[LVNMPY]-K-[STT]A-E-[PA]-[KSTT]A-[STT]AM.
NAME: PTD HPS conjugated serine phosphorylation site signature.
CONSENSUS: [KLSA]E[LVNMPY]-[KTH]-[KSA]H-[LVNMPY]D-[LVNMPY]-K[TH]-[LVNMPY]-[KSA]E.
NAME: PTD EEA domain phosphorylation site signature 1.
CONSENSUS: G-K[TH]-[LVNMPY]D-K-[LVNMPY]-G-[LVNMPY]-T-[LA]YV.
NAME: PTD EEA domain phosphorylation site signature 2.
CONSENSUS: [KSTTQ]-K[TH]-[LVNMPY]-[KSA]-K[TH]-[LVNMPY]-[LA]YV-K[TH]-[KSA]E.
NAME: PTD EER domain tyrosine phosphorylation site signature.
CONSENSUS: N-[LVNMPY]-K[TH]-C-K-T-S-[LVNMPY]-[LVNMPY]-[LVNMPY]-[KTH]-[KSA]E.
NAME: Adenylylase kinase signature.
CONSENSUS: [LVNMPY]Y[STT]-D-G-[PTV]D-P-K-K[TH]-[KMS].
NAME: Multimeric diaphosphorylation kinase active site.
CONSENSUS: N-K[TH]-N-[GSA]-S-D-[KSA]-[LVNMPY]K[TH].
NAME: Chaperone kinase signature.
CONSENSUS: T-[STT]-E-K[TH]-[KLS]-K[TH]-[KTH]-K[TH]-D-K[TH]-T-K-[PTT]-[LVNMPY]K[TH].

PCT/1300/01496

1046

PCT/IB02/01496

1047

PCT/IB00/01496

1048

NAME: Densinucleoside I sequence 1.
CONSENSUS: (LVMKZ)AAP(LR(STAKZ)P-AL)R(LVMHIDH)+L+(DE)V.
CONSENSUS: G-D-P-A-A-C-GAA.
NAME: Densinucleoside I sequence 2.
CONSENSUS: G-D-P-A-A-C-GAA.
NAME: Eubacterium 83 iron-sulfur binding region sequence.
CONSENSUS: C-KTH(EKZ)P(RAAGL)C-KTHC-KTHC.
NAME: Eubacterium 83 family segment.
CONSENSUS: (GTT)+L(VMP)P-KD(LPMW)K(LD(PAS)G V(GA)KTH(GAC)
CONSENSUS: K3(LPM)K3(EALV)L(VMP)W(GAR).
NAME: Eubacterium 8 family segment.
CONSENSUS: (HGTPTD)GTTA(LG(LPM)H-3)Y-ETALL-(P(VVAC)(TV)SA)P-L(VMA)
CONSENSUS: (HGTPTD)PTT+D-KTH(G)
NAME: Eubacterium 9 family segment.
CONSENSUS: (DEQ)TRQ(LAQ)R(PYV)LVY-G-D-BAK.
NAME: Internal Eubacterium 9 protein sequence segment.
CONSENSUS: (LPMPTB)K-D-K-D-R-PRD(SLQ)(LPM)(RBA)R+(LVMKA)(ER)
NAME: Eubacterium T2 family heptamer active site 1.
CONSENSUS: (PVM)K+L(VM)H-G-L-W-P.
NAME: Eubacterium T2 family heptamer active site 2.
CONSENSUS: (LPM)P-KD-PRDCTT(KD)P(W)+ (R)H-G-C.
NAME: Pseudomonas fluorescens family segment.
CONSENSUS: C-R-KTH-M-T-P.
NAME: DNA/RNA non-specific endonuclease active site.
CONSENSUS: D-R-G-R(GSL)K2A.
NAME: Thermotoga family segment 1.
CONSENSUS: D-G-Q-T(LVM)+L(VM)K-KH-R(LVM)K(LVM)D+P-R.
NAME: Thermotoga family segment 2.
CONSENSUS: D(LK)Y-GQK-R+(LVY)KA3+(VY)P(W)
NAME: Bacteriophage active site 1.
CONSENSUS: H-K-C-G-G-A-Y-G-D.
NAME: Bacteriophage active site 2.
CONSENSUS: G+(SA)G-S(LVM)R-T-P-S-Y.
NAME: Chlamydomonas reinhardtii active site segment.
CONSENSUS: (STH)G(LV)K(LD)(K)+W-E-S-KTH(GH).
NAME: Polypeptidase active site.
CONSENSUS: (SEDEH)K3-KTH-VMPK-KD(GH)H-G(LVMAK)K(LD)L(VM)G-S.
NAME: Chlamydomonas reinhardtii active site segment.
CONSENSUS: D(LVMP)T(GH)+ (DWD)KTH(LVM)H(SALM)K-D-KD(LVM)P+
CONSENSUS: (KZ)+L(VMP)
NAME: Chlamydomonas reinhardtii active site.
CONSENSUS: (LVMPT)H(KH)G(LVM)H(DM)L(VMP)H(DM)+S.
NAME: Chlamydomonas reinhardtii active site 1.
CONSENSUS: C-K-L-P-T-ETT-KTHPT(LVM)P+A-KTH(TP)K2-F(GSA).
NAME: Chlamydomonas reinhardtii active site 2.
CONSENSUS: (LVM)GSA-F+GTAGD(LVMP)T-W-PTT-W(LVM).
NAME: Alpha-helical protein / lysozyme C sequence.
CONSENSUS: C-KTHC-KD(LMP)K2(DSDH)SLD-KTHC.
NAME: Alpha-helical protein sequence.
CONSENSUS: G(LVMP)K3-K3(LVMP)T+(LVM)D-D-D+W-K3+R-DKSP.
NAME: Tyrosine sequence 1.

PCT/1303/01496

1050

1051

PCT/IB00/01496

1052

1053

WQ 01/12659 PCT/LB0601496

[illegible]

PCT/IB2006/01496

PCT/JP2004/01496

1056

CONSENSUS: [C]A[4]L[L]V[M]G[L]V[M]P[F]T[3]K[2]G[3]T[3]T[3]G[3]G[3]G[3]A[3]P[F]K[4]N[3]E[3]Q[3].
NAME: Dihydrodipicolinate synthetase segment 2.
CONSENSUS: Y[3]K[3]G[3]L[L]V[M]P[F]K[2]G[3]T[3]K[3]L[L]V[M]G[3]L[3]I[3]I[3]L[L]V[M]G[3]G[3]A[3]L[L]V[M]P[F].
CONSENSUS: E[3]D[3]E[3]A[3]P[F]T[3]A[3]C[3].
NAME: RNAi family of penicillinase synthase segment.
CONSENSUS: G[3]S[3]L[3]G[3]K[3]L[3]T[3]G[3]G[3]L[L]V[M]P[F]K[2]G[3]T[3]D[3]H[3]T[3].
NAME: Cytochrome synthetase/penicillinase-like cytochrome P-450-like monooxygenase site.
CONSENSUS: E[3]K[3]E[3]L[3]P[3]A[3]L[3]T[3]A[3]G[3]L[3]L[L]V[M]P[F]K[2]G[3]A[3]L[3]C[3]L[3]P[3]T[3]P[3]L[L]V[M]P[F]G[3]A[3].
NAME: Penicillinase and tetracycline resistance synthase segment.
CONSENSUS: G[3]T[3]G[3]L[L]V[M]G[3]T[3]G[3]L[3]C[3]S[3]G[3]D[3]H[3]L[3]P[3]L[3]G[3]A[3]K[3]L[3]I[3]A[3].
NAME: Penicillinase dihydrodipicolinate binding site.
CONSENSUS: E[3]K[3]K[3]L[L]V[M]P[F]K[2]G[3]T[3]L[L]V[M]P[F]K[2]G[3]A[3]L[3]C[3]L[3]P[3]T[3]P[3]L[L]V[M]P[F]G[3]A[3].
NAME: Cytochrome monooxygenase synthetase/penicillinase monooxygenase site.
CONSENSUS: [D]Q[3]L[L]V[M]P[F]K[2]G[3]T[3]A[3]G[3]L[3]T[3]A[3]G[3]T[3]T[3]G[3]P[3]Y[3]W[3]L[L]V[M]P[F]K[2]G[3]K[3]Q[3]S[3]K[3]H[3]K[3].
NAME: Chitinase I segment 1.
CONSENSUS: [H]Q[3]L[L]V[M]P[F]K[2]G[3]T[3]P[F]T[3]K[3]L[3]P[3]T[3]K[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F].
NAME: Chitinase I segment 2.
CONSENSUS: G[3]P[3]T[3]G[3]K[3]L[3]H[3]L[3]G[3]A[3]L[3]V[P]F[3]K[3]H[3]K[3]P[3]V[P]K[3]G[3]A[3]L[3]T[3]A[3]G[3]L[3]I[3]E[3]W[3].
NAME: Cytochrome c and c1 haem fusion segment 1.
CONSENSUS: M[3]K[3]K[3]D[3]H[3]K[3]E[3]L[3]L[3]W[3]Q[3]K[3]E[3]K[3]L[3]W[3]E[3].
NAME: Cytochrome c and c1 haem fusion segment 2.
CONSENSUS: P[3]S[3]G[3]L[3]D[3]V[3].
NAME: Adenylate cyclase-like domain segment 1.
CONSENSUS: S[3]T[3]P[3]G[3]S[3]A[3]T[3]L[3]W[3]L[3]T[3]E[3].
NAME: Adenylate cyclase domain segment 2.
CONSENSUS: T[3]K[3]P[3]K[3]W[3]Q[3]L[3]L[L]V[M]G[3]L[3]T[3]L[3]E[3]P[3]G[3].
NAME: Chitinase I segment.
CONSENSUS: D[3]V[3]L[L]V[M]G[3]K[3]L[3]G[3]K[3]L[3]P[3]T[3]K[3]L[3]V[M]G[3]P[3]Y[3]W[3]Q[3]L[3]W[3]T[3]K[3]W[3]Q[3]H[3]T[3]P[3]V[3].
CONSENSUS: [D]H[3]T[3]A[3]K[3]D[3]E[3].
NAME: Chitinase I segment 1.
CONSENSUS: G[3]E[3]S[3]H[3]K[3]K[3]L[3]L[3]V[M]G[3]K[3]T[3]V[3]L[3]L[L]V[M]G[3]T[3]D[3]E[3]G[3]K[3]P[3]V[3].
NAME: Chitinase I segment 2.
CONSENSUS: [Q]E[3]K[3]G[3]A[3]L[3]T[3]A[3]G[3]K[3]P[3]V[3]L[3]T[3]K[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F]K[2]G[3]K[3]L[3]I[3]A[3].
NAME: Chitinase I segment 3.
CONSENSUS: S[3]E[3]H[3]G[3]P[3]Y[3]K[3]L[3]A[3]V[3]K[3]L[3]I[3]G[3]A[3]D[3]L[3]V[3]G[3]P[3]L[L]V[M]G[3]L[3]T[3]A[3]G[3]L[3]I[3]A[3].
NAME: 6-pyruvoyl tetrahydrophosphate synthase segment 1.
CONSENSUS: C[3]K[3]K[3]K[3]D[3]G[3]N[3]G[3]H[3]N[3]Y[3].
NAME: 6-pyruvoyl tetrahydrophosphate synthase segment 2.
CONSENSUS: D[3]H[3]K[3]K[3]L[3]I[3]L[3]D[3]L[3]I[3]D[3].
NAME: Penicillinase segment.
CONSENSUS: [L]V[M]P[F]T[3]K[2]G[3]L[3]K[3]G[3]S[3]L[L]V[M]G[3]P[3]K[3]L[3]I[3]D[3]E[3]K[3]Q[3]K[3]K[3]G[3]D[3]P[3]V[3].
NAME: Abietic monooxygenase synthetase/penicillinase monooxygenase site.
CONSENSUS: Y[3]K[3]E[3]K[3]G[3]P[3]L[3]G[3]A[3]L[3]T[3]A[3]G[3]L[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F].
NAME: Arginine and glutamine monooxygenase segment 1.
CONSENSUS: [P]V[3]L[L]V[M]G[3]K[2]G[3]T[3]K[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F]T[3]A[3]N[3]K[3].
NAME: Arginine and glutamine monooxygenase segment 2.
CONSENSUS: [L]V[M]G[3]T[3]A[3]G[3]L[3]T[3]A[3]G[3]L[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F]T[3]A[3]N[3]K[3]L[3]L[L]V[M]G[3].
NAME: Mandelate racemase / mandelate lyase/mandelate lyase family segment 1.
CONSENSUS: A[3]K[3]L[3]A[3]C[3]T[3]L[L]V[M]G[3]T[3]D[3]E[3]K[3]L[3]I[3]A[3]G[3]L[3]L[3]I[3]L[3]V[M]G[3]L[3]M[3]P[F]T[3]A[3]N[3]K[3].
NAME: Mandelate racemase / mandelate lyase/mandelate lyase family segment 2.

1058

NAME: Aminoacyl-tRNA synthetase class-1 signature
CONSENSUS: P-AE-DGTTAAQ(DENQGLSPK)+ELVMPF(RTH)(LVMYAC)G-SHWTCQ
CONSENSUS: [LVMPTETADPC]

NAME: Aminoacyl-tRNA synthetase class-8 signature 1
CONSENSUS: (PTP)E+G(DG+L)EAE+K(P+K)DDE

NAME: Aminoacyl-tRNA synthetase class-8 signature 2
CONSENSUS: (GTAL)YF(DENQGLSPF)(GTAA)(LVMPL(DG)E(LVMF)+ELVMTAG)(LVMFP)

NAME: WHEP-TS domain signature
CONSENSUS: (QTV)G(DHKA)+ELVY(DRE)K+K(LKNG)(AD)H+ELVY(DENK)
CONSENSUS: K(DVY)K(L-L)G(E

NAME: ATP-cases/leucyl-tRNA synthetase family signature 1
CONSENSUS: S(LR)A-G(DT)(LVM)(KSTT)+(DQ)K(LR)G+K(LVM)(DA)(LVM)+G
CONSENSUS: G-D

NAME: ATP-cases/leucyl-tRNA synthetase family active site
CONSENSUS: G-KT+A+K(L)H(DT)(LVMF)G-A(AE)(GL)

NAME: ATP-cases/leucyl-tRNA synthetase family signature 2
CONSENSUS: G+(PT)K(LVMF)+DA(GDA)G(LA)(STAV)K(D+LVM)+K
CONSENSUS: G(KDE)

NAME: Chlamydomonas signature 1
CONSENSUS: (PTW)L(D-G-S-E)K(DENQGTAR)(SA)(DE)+K(LVMFP)

NAME: Chlamydomonas putative ATP-binding region signature
CONSENSUS: E-P(LVMFPY)A-K(L)H(MFA)G-DGTAM)G+H-K(E

NAME: Chlamydomonas class-1 subfamily site
CONSENSUS: S(LVM)K(D)ELVMA(DHSD)(DH)(L)H-Y

NAME: D-ribitol-D-ribitol ligase signature 1
CONSENSUS: H-G-K(D)G-E-D-G+ELVMAA(GSA)(SEA)

NAME: D-ribitol-D-ribitol ligase signature 2
CONSENSUS: ILTY+K(D)GSA+KCAVYR(LVCA)D(LVMF)K(D)K(L)H+K
CONSENSUS: (L)Y(A)H(DTP)+P(KA)

NAME: SAKCAR signature signature 1
CONSENSUS: (LVMF)D-P(LVM)K+ELVMA(LVMCA)R-K(LTA)G-E

NAME: SAKCAR signature signature 2
CONSENSUS: (LVM)(LVM)A-D+K(LVMFP)E-P-G

NAME: Polyphosphatase signature 1
CONSENSUS: (LVMFP)+ELVMA(DT)A(Q)G-T(PK)G-E+(ST)K(LVM)D(KHSD)

NAME: Polyphosphatase signature signature 2
CONSENSUS: (LVMFPY)D(E+G(LVM)KA)G-K(D)G+KSTT+(LVM)D

NAME: Unspecific-acting enzyme signature 1
CONSENSUS: E-A-C-S-G-S-P+P

NAME: Unspecific-acting enzyme active site
CONSENSUS: P(LVM)C(T-LVM)(L)H+(PT)P

NAME: Unspecific-acting enzyme active site
CONSENSUS: (PTW)LDPH-K(L)H(DH)(L)H-K(L)H+G+ELVY-C(LVY)+(L)H

NAME: Putative-oxidoreductase signature 1
CONSENSUS: Q(LVM)S-G-G-A-L-L-G-D-Y

NAME: Putative-oxidoreductase signature 2
CONSENSUS: V-A-T(DV)E-A-L-E+(HW)G-G

NAME: Adenosine/adenosine signature (GTP-binding site)
CONSENSUS: Q-W-G-D-S-G-E-G

NAME: Adenosine/adenosine signature active site
CONSENSUS: G(L)G(R)P+T-K(D)E-K(D)R

NAME: Argininosuccinate synthase signature 1.
CONSENSUS: A-P-T-I-S-Q-L-D-T-S
NAME: Argininosuccinate synthase signature 2.
CONSENSUS: G-T-T-E-G-N-Q-K-T-E-F
NAME: Phosphoenolpyruvate-carboxy synthase signature.
CONSENSUS: R-R-Q-D-E-P-Q-M
NAME: Carboxymethyl-phosphate synthase subfamily signature 1.
CONSENSUS: (P-T-Y-Y)(P-F-S-L-Y-N-E-Z)(L-Y-N-M-A)(L-Y-N-M)(E-K-I)(P-S-A)(D-T-A)-K-T-(P-K-D-G-(A-G)
NAME: Carboxymethyl-phosphate synthase subfamily signature 2.
CONSENSUS: (L-Y-N-M)(L-S-N-P-A-E-L-Y-N-M-A)-R-(P-A-T-L-Y-N-M)(E-K-I)(L-Y-N-M-T-A-C)
NAME: ATP-dependent DNA ligase AMP-binding site
CONSENSUS: (E-D-Q-N)+R+L-D-N-Y-G+R-(G-A-C-T-Y-M)
NAME: ATP-dependent DNA ligase signature 2.
CONSENSUS: R-G-L-I-Y-M-A-I-L-Y-N-E-Z-(E-K-I)-K-I-T-Y-W-H-Q-E-K-I-K-I-L-A-I-E-N-I-K-I-D-I-E-
CONSENSUS: (L-Y-N-M-T-Y-E
NAME: NAD-dependent DNA ligase signature 1.
CONSENSUS: K-L-Y-N-M-D-D-G-L-Y-N-M-(E-K-I)-K-I-Y-K-T-D-D-I-L-M-I-D-T-T-R-Q-D-(D-M)-Q-K-T-D-G-
CONSENSUS: (D-S-I-D-S-I-L)
NAME: NAD-dependent DNA ligase signature 2.
CONSENSUS: (P-Y)-G-(E-K-I)-D-T-T-G+(L-Y-N-M)(D-T-N-E-I)+(P-T)-K-D-L-(P-S)-Y.
NAME: RNA 3' terminal phosphate cyclase signature.
CONSENSUS: (H-M-I-G-K-D)-P+(Q-D)-L-L-Y-I
NAME: Ligase protein ligase II signature.
CONSENSUS: R-G-Q-K-D-T-P-P-Y-W-I-E-K-T-E-K-I-Q-(L-Y-I)+Y.
NAME: Inositol-like N synthase signature 1.
CONSENSUS: (R-K-I)+(D-T-A)-K-D-S-A-C-T-C-D-L
NAME: Inositol-like N synthase signature 2.
CONSENSUS: (L-Y-N-M)(D)-C-G-(D-T-A)-K-D-(D-T-A-G)-K-T-T-(D-M-G)
NAME: Site-specific recombinase active site
CONSENSUS: Y-L-Y-A-C-R-Y-A-I-S-I-T-T-K-D-L-Q
NAME: Site-specific recombinase signature 2.
CONSENSUS: G-(D-K)-K-D-L-Y-N-M-K-T-(L-Y-N-M)(D-T-T)-L-(L-Y-N-M)(E-K-I)
NAME: Transposase, hobo family, signature.
CONSENSUS: D-K-T-G-L-Y-N-M-Y-K-M-D-T-A-T-Y-(L-Y-N-M-T-Y-W)-(P-T)+(D-T-A-Y)-K-T-(E-K-I)+C-K-T-
CONSENSUS: H
NAME: Transposase, IS30 family, signature.
CONSENSUS: R-G-K-T-D-R-M+R-G-L-Y-N-M-D-R-(E-K-I)(L-Y-N-M-Y-K-T)-F-E
NAME: Apurinic/apyrimidinic family signature.
CONSENSUS: (L-M-F-T-I-S-K-D-F-K-T-E-K-I)-W+(L-Y-N-M)(K-A-9)-E+D+L-(P-T)-D
NAME: Thiamine pyrophosphatase enzyme signature.
CONSENSUS: (L-Y-N-M)(E-S-A)-K-D-F-K-I-Q-L-Y-N-M-T-Y-W)-(L-Y-N-M)+G-D-K-D-A-I-(G-A-C)
NAME: Biotin-regulating enzyme subfamily site
CONSENSUS: (D-M-I-D-E-Q-T-T-G)+(L-Y-N-M-T-T)-K-T-L-Y-N-M+(A-P-Y)-M-E-(L-M-A-T)-K-T-L-Y-N-M+
CONSENSUS: (S-A-Y)
NAME: Zinc and dithioloprotein cysteine/thioloprotein cysteine type binding site
CONSENSUS: (D-M)-K-D-L-Y-T-Y-K-T-D-L-Y-N-M-K-D-L-Y-N-M-A-K-T-L-D-T-A-Y-Y-D-T-A-Y-D-M-G
CONSENSUS: K-D-L-Y-N-M-T-E-K-T-(E-K-I)-L-Y-N-M-T-Y
NAME: Putative AMP-binding domain signature.
CONSENSUS: (L-Y-N-M-T-Y)-K-T-(D-T-G)(D-T-A-G)-Q-(D-T-T)(D-T-E-K)(G-C)+(P-A-S-L-Y-N-M)(K-E)
NAME: Methyltransferase sulfonate transferase protein signature 1.
CONSENSUS: (L-Y-N-M)(S-L-T-T-E-D-G-D-Y-D-K-A-D)

NAME: Multidomain substrate hexapeptide protease signature 2.
CONSENSUS: S+(GDS-KD-D-KD)(LNVW)-K(R, L)(LNV) -K(D)(R, S)+G(DRL)P-K-D.
CONSENSUS: (LNVWPT)KGL

NAME: BmsA / mB / pBst family signature.
CONSENSUS: (LTV)K(D)C-(PTV)(LNVWPT)QKSG-C-(PTM)C.

NAME: Radical scavenging enzyme signature.
CONSENSUS: (GV) + G + (SR) -K(D)-P-K(D)-G -K(D)-C-K(D)-C-K(D)-C + (H).
CONSENSUS: S-K-D-L-P-P-A-K(D)(R, S)(PTW)-C.

NAME: Cyclohexane 1 family beta-binding site signature.
CONSENSUS: C-(PWPWP)-Z(PWZ)-C-H-K(PYW).

NAME: Cyclohexane 1 family, beta-binding domain signature.
CONSENSUS: (PTV)(LNVW)K-L-K-D-H-P-K(D)G.

NAME: Cyclohexane 1b family beta-binding site signature.
CONSENSUS: (DSD)K -K(D)-G-(PTW)WQD-(LNVWPT)S-K(D)H.

NAME: Cyclohexane 1bA Qa site signature.
CONSENSUS: P-(DQ)-P-(PTV)(LPT)G.

NAME: Cyclohexane 1bB substrate beta-binding site signature.
CONSENSUS: (LTV) + (ST)(LTVV)S-(PTV)-K(D)-H(V)-H-ST(DAL-LV)-ST(DAL)H(V)P.

NAME: Nickel dependent hydrogenase b-type cyclohexane subunit signature 1.
CONSENSUS: S-L(LNVWPTW)+H-W-(LNVW) -K(D)(LNVWPT)ST-K(LNVW)-K(D)-L-L(LNVW)-T-G.

NAME: Nickel dependent hydrogenase b-type cyclohexane subunit signature 2.
CONSENSUS: (H)G-ST(A)(LNVWPTW)+H-L(K)(LNVW)-K(D)-W-(LNVWPT) -K(D)-P-K(D)-H.

NAME: Succinate dehydrogenase cyclohexane b subunit signature 1.
CONSENSUS: S-P-L(PVW) -K(D)(LNVW)-K(D)(LNVW)WPK-L-K(D)-S-K(D)-H-S + (ST).

NAME: Succinate dehydrogenase cyclohexane b subunit signature 2.
CONSENSUS: H-K(D)K(A)(LNVWPT)S-H(H)K(LTVW)-L-(PTW)W-K(D)-D + (QVA).

NAME: Thermophilic family subunit site.
CONSENSUS: (LNVW)S-L(PWST)A-L-L(LNVWPT)PTWWTW(K) -K(D)-PTWWTW(K)-C-K(A)PLV(K)-
(PW)WTWTA(C)-K(LNVWPTW).

NAME: Osmotolerant family site.
CONSENSUS: (LNV)S-(PTA)S-K(L)-C-(PTV)(PTW)-C-K(D)(TAV) -K(D)-H-L(V).

NAME: Type 1 super (bim) protein signature.
CONSENSUS: (K)A-K(L)H-T(A)K-H-L(LVPT)+C-K-K(L)Z(PQ) -K(D)-H-K-K(D)-H-D(KQ).

NAME: 2P-18 ferredoxin, iron-sulfur binding region signature.
CONSENSUS: C-(P)-K(C)-K(A)-K(C)-C-(K)A(TT)-(CPKELNPTW)-C.

NAME: Adenosine family, iron-sulfur binding region signature.
CONSENSUS: C-K(D)-ST(A)S-K(TA)WPT-C-(TTA)T-C-(H).

NAME: 4P-45 ferredoxin, iron-sulfur binding region signature.
CONSENSUS: C-K(D)-C-K(D)-C-K(D)-C-(P)KQ.

NAME: High potential iron-sulfur protein signature.
CONSENSUS: C-K(L)S-(LNVW) -K(D)-G-(TV)-C-K(D)-H-L(V).

NAME: RndA iron-sulfur protein signature 1.
CONSENSUS: C-(TA)K-L-G-C-L(V).

NAME: RndA iron-sulfur protein signature 2.
CONSENSUS: C-P-C-H + (QSA).

NAME: Phenolase signature.
CONSENSUS: (LTV)(LNVV) (PTV) + (ST) -K(D)-A(KC) + T-K(H) -K(D)(LNV).

NAME: Reductase signature.
CONSENSUS: (LNVW) -K(D)-W-K-C-P-K-C-(AGQ).

1062

PCT/US00/01496

1063

[illegible]

WO 81/12659

PCT/IB80/01796

NAME: General difference Chain-sequence protein sequence.
CONSENSUS: [L]V[M]P[Y]-K-D-G-K-D-Y + P + R-K-D-T-H-S-H-V-[H]T[V]-L-L-V-M-P-Y-W-Y.

NAME: Onopeltide domain.
CONSENSUS: [L]V[M]A-I + [G]T + (T)A-(D)A-K-D-[D]G-[G]E[Y]-K-T-[L]V-Y-D-E-[H]G-K-D-
CONSENSUS: [L]D-[D]S-[D]S-[E]R[D]S-K-A-K-D-[L]V-Y-K-D-[L]V-M-P-Y-H-I-J-[L]V-M-K-D-
CONSENSUS: [L]V-M-K-D-[D]S-K-D.

NAME: Poliovirus nonstructural protein sequence.
CONSENSUS: [T]H-K-D-D-[D]P-A-I + [D]T-A-I-K-D-[T]A-G-[E]K-I-[L]V-M-P-[D]H-T-A-I-[D]H-G-K-D-
CONSENSUS: [D]T-A-I-G-[E]P-H-A-I-I-[L]V-M-T.

NAME: Insulin-like (p)-A factor binding protein sequence.
CONSENSUS: G-C-[D]S-C-C-K-D-Y-C-A-K-H-C.

NAME: GP63/PLN30/ant body sequence.
CONSENSUS: H-P-A-Y-T-P-[L]V-P-C-L + [E]S-A-I-P.

NAME: GHKL1/SL24 family sequence.
CONSENSUS: L-I + P-L-A-I-Y-H-H.

NAME: C3 EA protein/mucin protein sequence.
CONSENSUS: Q-Q-D-Q-T-S-Q-Q-L.

NAME: Antigen sequence 1.
CONSENSUS: [P]T-[L]V-Y-G-[D]E-S-A-Q + [R]A-Q-E-I-Q.

NAME: Antigen sequence 2.
CONSENSUS: W-[D]T-[D]T-A-[E]K-I + [D]E-I-Y-[D]H-Q-[D]E-L.

NAME: Antigen and non-antigen protein sequence.
CONSENSUS: [L]M-Q-[L]V-M-Q-T-S-[D]A-P-Q-D + [L]V-M-P-Y-W-Q-H-[P]E-T-A-Q-K-D-H-[E]K-I.

NAME: Antigenic repeated domain sequence.
CONSENSUS: [T]H-[E]T-T-Y-K-H-[L]V-M-P-K-D-S-K-D-[D]E-Q-W-Y-K-D-[D]T-Y-K-D-[L]V-M-P-
CONSENSUS: K-D-[L]V-M-P-K-I-I-[L]V-M-P-A-K-D-[L]V-M-P.

NAME: Cysteine sequence.
CONSENSUS: P-S-D-Y-A-A-S-P.

NAME: Cholesterol high chain sequence 1.
CONSENSUS: P-L-A-Q-Q-S-S.

NAME: Cholesterol high chain sequence 2.
CONSENSUS: [E]K-I-D + S-[E]K-I-[L]V-M-Q-[E]K-I + [L]V-M-X-D-Y + L-K.

NAME: Cholesterol sequence 1.
CONSENSUS: C-E-P-C-L-E + T-C.

NAME: Cholesterol sequence 2.
CONSENSUS: C-L-[E]K-I-H-[E]K-I + [D]E-C-[E]D-S-K-C.

NAME: Cholesterol sequence 3.
CONSENSUS: C-D-H-I-Y + Q-P-Q-C-K-D-Y-C-Y-D.

NAME: Cholesterol sequence 4.
CONSENSUS: C-L-I-I + P-C-K-D-[L]V-M-Q-[D]H-Q-C-[P]T-[L]V-M-Q-[S]A-[E]K-I-P.

NAME: Cysteine rich and protein "linker" motif sequence.
CONSENSUS: [L]V-M-P-Y-A-I-I + [D]D-S-L-E-T-T-Y-[P]T-I-D-H-S-L-Y-T-I-K-D-[P]T-Y + G-K-D-[L]V-M-P-K-D-T-T.

NAME: Dynamic family sequence.
CONSENSUS: L-P-[E]K-I-G-[E]T-T-Y-[E]K-I-[L]V-M-Q-Y-T-S.

NAME: Dynamic high chain type 1 sequence.
CONSENSUS: W-K-I + G-[D]S-I + P-[D]S-A-I-S + V-[D]T-[D]T-Y-S.

NAME: Phe2 protein sequence 1.
CONSENSUS: H-[D]T-D-S-Q + L-K-I-L-I-D-G + G-I-A-T-Y-G-[D]S-A-P + P-K-D-D.

NAME: Phe2 protein sequence 2.
CONSENSUS: [D]H-S-I-[L]V-M-P + [L]V-M-P-Q-D-[V]T-A-C-[E]T-A-C-G + G-[D]S-I-G-T-D-[D]T-Y-G-
CONSENSUS: [D]S-A-I-[D]T-A-I-P-[L]V-M-P-T-[L]V-M-P-[D]S-A-I-Y.

1066

[illegible]

1068

PCT/BO2014/01494

1069

1070

PCT/IB02/01496

1071

1072

CONSENSUS: $(LVM^T)^3$
 NAME: Chapman TCF 1 equation 1.
 CONSENSUS: $(B E L E T T) + (L M P T) \cdot G + (G S A) \cdot + + E (L V M P E D)$
 NAME: Chapman TCF 1 equation 2.
 CONSENSUS: $(L V M H T B T S M L S D E S A) (L V M H S L T A V) (L V M H S D L D L P M) + (L V M) +$
 CONSENSUS: $(B M H) (P Q M)$
 NAME: Chapman TCF 1 equation 3.
 CONSENSUS: $Q (D E S A) \cdot + + (L V M E T T A) (S A) \cdot O \cdot T$
 NAME: How much log/76 protein family profile.
 NAME: How much log/76 protein family equation 1.
 CONSENSUS: $(P V) \cdot O \cdot L \cdot O \cdot T (S T) + (S R)$
 NAME: How much log/76 protein family equation 2.
 CONSENSUS: $(L V M P T L V M P T T) (D P) (L V M P E D) \cdot Q (G S A) (S T) (A T T) (D T) (L V M)$
 CONSENSUS: $(L V M P C)$
 NAME: How much log/76 protein family equation 3.
 CONSENSUS: $(L V M T) + (L V M P T) + O \cdot O + (S T) + (L V M) \cdot P + (L V M) + (D E G E R T A)$
 NAME: How much log/76 protein family equation.
 CONSENSUS: $T + (H Q S) E (S E) (P A L F L L E D E)$
 NAME: Chapman (qpa/v) equation 1.
 CONSENSUS: $D (A S) (S A) M (L V M P E D) H (P T) + L \cdot K \cdot D$
 NAME: Chapman (qpa/v) equation 2.
 CONSENSUS: $E (L V M P T T) D + S E (L V M P T) + S (S A Q) + (S T A) + (S T A) (K S) (L V M) + O$
 CONSENSUS: $(S T A)$
 NAME: H-dm/ distance equation.
 CONSENSUS: $(P T) \cdot K \cdot D (L V M A) \cdot K \cdot T (P V M H T) (D E G S A) + L + (D H) \cdot K \cdot T (S) \cdot K \cdot D (P T)$
 NAME: dm/ distance profile.
 NAME: CYC/CYC/ dm/ distance equation.
 CONSENSUS: $C (D E G E T S A) + C + G + (S A) (A C C D M) \cdot K \cdot D (S M H E) \cdot K \cdot A \cdot B \cdot C \cdot K \cdot D \cdot B \cdot C + O + G$
 NAME: gyl/ protein equation.
 CONSENSUS: $(P L) (D P T T E A) \cdot K \cdot D (S A) + + A (L V M T M) \cdot K \cdot H \cdot M \cdot O \cdot Q (P T) \cdot K \cdot T (D E G) \cdot K \cdot D$
 CONSENSUS: $(L V M) (S A) + (S A) + V + (P T)$
 NAME: Recurrent type II mutation system protein C equation.
 CONSENSUS: $P \cdot M \cdot P \cdot K \cdot D \cdot L \cdot K \cdot D (L V M A) \cdot (L V M) + (L V M) M + (L V M) + L$
 NAME: Recurrent type II mutation system protein D equation.
 CONSENSUS: $H S (D E G C) (S T T V M) (L V M A S) (S A) (S L V M P T T) (L V M) \cdot P$
 CONSENSUS: $(L V M P T T) (S L V M P T T) (S A) \cdot (L V M) \cdot P (L V M P T T) (S L V M P T T) \cdot K \cdot D (S L V M P T T)$
 NAME: Recurrent type II mutation system protein E equation.
 CONSENSUS: $(L V M) \cdot K \cdot D \cdot P \cdot D + (L V M) \cdot D \cdot S (L V M) \cdot S \cdot D$
 NAME: Recurrent type II mutation system protein F equation.
 CONSENSUS: $(S A Q) (L V M A) \cdot K \cdot D (S A) (L V M) + (T T T) \cdot P \cdot K \cdot D (L V M) \cdot K \cdot D (S T A C T) \cdot K \cdot D$
 CONSENSUS: $(L M T) \cdot K \cdot D (L V M P T T) \cdot P$
 NAME: Recurrent type II mutation system protein H equation.
 CONSENSUS: $O \cdot T \cdot L \cdot M \cdot O \cdot K \cdot D (S A) \cdot M$
 NAME: Recurrent type II mutation system protein I equation.
 CONSENSUS: $S (L V M) (S A) \cdot S \cdot Y (S A) \cdot A \cdot S \cdot P (S T V M) \cdot D (S A) \cdot M \cdot P \cdot G \cdot S \cdot Q \cdot M (S A) \cdot D$
 CONSENSUS: $(S A) \cdot O$
 NAME: Protein mut equation.
 CONSENSUS: $(P T) + (P T) (S A) \cdot T (P Q) \cdot M \cdot A \cdot G \cdot O + O + L$
 NAME: Protein mut equation 1.
 CONSENSUS: $(S E T) (L V M P T T) \cdot (L V M) \cdot O (L V M) + P (L V M P T T) + (A S) (S T T)$
 CONSENSUS: $(L V M P T T) \cdot D (L V M P T T)$

1074

WO 01/12659 PCT/JP00/01496

```

NAME: Arbitrary segment
CONSERVED: 178-179-180-181-182-183-184-185-186-187-188-189-190-191-192-193-194-195-196-197-198-199-200-201-202-203-204-205-206-207-208-209-210-211-212-213-214-215-216-217-218-219-220-221-222-223-224-225-226-227-228-229-230-231-232-233-234-235-236-237-238-239-240-241-242-243-244-245-246-247-248-249-250-251-252-253-254-255-256-257-258-259-260-261-262-263-264-265-266-267-268-269-270-271-272-273-274-275-276-277-278-279-280-281-282-283-284-285-286-287-288-289-290-291-292-293-294-295-296-297-298-299-300-301-302-303-304-305-306-307-308-309-310-311-312-313-314-315-316-317-318-319-320-321-322-323-324-325-326-327-328-329-330-331-332-333-334-335-336-337-338-339-340-341-342-343-344-345-346-347-348-349-350-351-352-353-354-355-356-357-358-359-360-361-362-363-364-365-366-367-368-369-370-371-372-373-374-375-376-377-378-379-380-381-382-383-384-385-386-387-388-389-390-391-392-393-394-395-396-397-398-399-400-401-402-403-404-405-406-407-408-409-410-411-412-413-414-415-416-417-418-419-420-421-422-423-424-425-426-427-428-429-430-431-432-433-434-435-436-437-438-439-440-441-442-443-444-445-446-447-448-449-450-451-452-453-454-455-456-457-458-459-460-461-462-463-464-465-466-467-468-469-470-471-472-473-474-475-476-477-478-479-480-481-482-483-484-485-486-487-488-489-490-491-492-493-494-495-496-497-498-499-500-501-502-503-504-505-506-507-508-509-510-511-512-513-514-515-516-517-518-519-520-521-522-523-524-525-526-527-528-529-530-531-532-533-534-535-536-537-538-539-540-541-542-543-544-545-546-547-548-549-550-551-552-553-554-555-556-557-558-559-560-561-562-563-564-565-566-567-568-569-570-571-572-573-574-575-576-577-578-579-580-581-582-583-584-585-586-587-588-589-590-591-592-593-594-595-596-597-598-599-600-601-602-603-604-605-606-607-608-609-610-611-612-613-614-615-616-617-618-619-620-621-622-623-624-625-626-627-628-629-630-631-632-633-634-635-636-637-638-639-640-641-642-643-644-645-646-647-648-649-650-651-652-653-654-655-656-657-658-659-660-661-662-663-664-665-666-667-668-669-670-671-672-673-674-675-676-677-678-679-680-681-682-683-684-685-686-687-688-689-690-691-692-693-694-695-696-697-698-699-700-701-702-703-704-705-706-707-708-709-710-711-712-713-714-715-716-717-718-719-720-721-722-723-724-725-726-727-728-729-730-731-732-733-734-735-736-737-738-739-740-741-742-743-744-745-746-747-748-749-750-751-752-753-754-755-756-757-758-759-760-761-762-763-764-765-766-767-768-769-770-771-772-773-774-775-776-777-778-779-780-781-782-783-784-785-786-787-788-789-790-791-792-793-794-795-796-797-798-799-800-801-802-803-804-805-806-807-808-809-810-811-812-813-814-815-816-817-818-819-820-821-822-823-824-825-826-827-828-829-830-831-832-833-834-835-836-837-838-839-840-841-842-843-844-845-846-847-848-849-850-851-852-853-854-855-856-857-858-859-860-861-862-863-864-865-866-867-868-869-870-871-872-873-874-875-876-877-878-879-880-881-882-883-884-885-886-887-888-889-890-891-892-893-894-895-896-897-898-899-900-901-902-903-904-905-906-907-908-909-910-911-912-913-914-915-916-917-918-919-920-921-922-923-924-925-926-927-928-929-930-931-932-933-934-935-936-937-938-939-940-941-942-943-944-945-946-947-948-949-950-951-952-953-954-955-956-957-958-959-960-961-962-963-964-965-966-967-968-969-970-971-972-973-974-975-976-977-978-979-980-981-982-983-984-985-986-987-988-989-990-991-992-993-994-995-996-997-998-999-1000
CONSERVED: 1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53-54-55-56-57-58-59-60-61-62-63-64-65-66-67-68-69-70-71-72-73-74-75-76-77-78-79-80-81-82-83-84-85-86-87-88-89-90-91-92-93-94-95-96-97-98-99-100
NAME: Uniquely determined segment
CONSERVED: 1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53-54-55-56-57-58-59-60-61-62-63-64-65-66-67-68-69-70-71-72-73-74-75-76-77-78-79-80-81-82-83-84-85-86-87-88-89-90-91-92-93-94-95-96-97-98-99-100
NAME: Uniquely determined segment
CONSERVED: 1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53-54-55-56-57-58-59-60-61-6
```

PCT/2000/014766

1076

[illegible]

PCT/IB00/01496

1078

1079

1080

[illegible]

What claim:

1. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16c16; hfr2_16f21; hfr2_16g18; hfr2_16i12; hfr2_16k22; hfr2_16l12; hfr2_22f21; hfr2_22h13; hfr2_22h13; hfr2_22i4; hfr2_22k3; hfr2_22k8; hfr2_22b10; hfr2_22b21; hfr2_22f2; hfr2_23i24; hfr2_23n16; hfr2_23c24; hfr2_23o5; hfr2_2a2; hfr2_2b17; hfr2_2b5; hfr2_2c1; hfr2_2c17; hfr2_2c18; hfr2_2d15; hfr2_2d17; hfr2_2d20; hfr2_2g18; hfr2_2h1; hfr2_2h10; hfr2_2i17; hfr2_2k14; hfr2_2k19; hfr2_2b16; hfr2_2c18; hfr2_2f16; hfr2_2g8; hfr2_2i2; hfr2_41m15; hfr2_62b11; hfr2_62f10; hfr2_62i19; hfr2_62n10; hfr2_62o17; hfr2_64a11; hfr2_64a15; hfr2_64c16; hfr2_64c4; hfr2_64b6; hfr2_64i20; hfr2_64j18; hfr2_64k24; hfr2_64o16; hfr2_6a17; hfr2_6b24; hfr2_6i20; hfr2_6o17; hfr2_71a20; hfr2_72b18; hfr2_72d13; hfr2_72i12; hfr2_72m16; hfr2_72n12; hfr2_78c24; hfr2_78d13; hfr2_78k24; hfr2_78a23; hfr2_7a24; hfr2_7c22; hfr2_7j4; hfr2_82c20; hfr2_10c20; hfr2_82a17; hfr2_82a4; hfr2_10a4; hfr2_82g14; hfr2_10g14; hfr2_82i17; hfr2_10; hfr2_82i24; hfr2_10; hfr2_82m16; hfr2_10; hfr2_82m6; hfr2_10; hfr2_1j9; hfr2_24a15; hfr2_24b15; hfr2_24c23; hfr2_24d20; hfr2_24p5; hfr2_3i13; hfr2_3o17; hfr2_46a6; hfr2_46b10; hfr2_46d13; hfr2_46j20; hfr2_46k19; hfr2_46m4; hfr2_47a4; hfr2_49b; hfr2_4c8; hfr2_4h14; hfr2_4m11; hncf1_1a11; hncf1_1c23; hncf1_1c15; hncf1_1g13; hncs3_1n3; hncs3_14g5; hncs3_14h21; hncs3_14p14; hncs3_14p7; hncs3_15a13; hncs3_15c24; hncs3_15c6; hncs3_15g14; hncs3_15h1; hncs3_15i5; hncs3_15j18; hncs3_15j3; hncs3_15k11; hncs3_17f10; hncs3_17f17; hncs3_17n12; hncs3_17n18; hncs3_18j3; hncs3_18j7; hncs3_19f19; hncs3_19j17; hncs3_1c1; hncs3_1g13; hncs3_1k11; hncs3_20c21; hncs3_20k2; hncs3_20m18; hncs3_21d4; hncs3_21j15; hncs3_21i16; hncs3_21n23; hncs3_22c23; hncs3_22g2; hncs3_22n13; hncs3_23i11; hncs3_23n19; hncs3_23o19; hncs3_26g22; hncs3_27d1; hncs3_27h4; hncs3_27o14; hncs3_28d14; hncs3_2a11; hncs3_2a17; hncs3_2d15; hncs3_2e12; hncs3_2f14; hncs3_2g7; hncs3_2h1; hncs3_2h15; hncs3_2i19; hncs3_2m18; hncs3_2m20; hncs3_2n9; hncs3_2o43; hncs3_30f4; hncs3_33b4; hncs3_33b5; hncs3_35c21; hncs3_35g6; hncs3_35k16; hncs3_35k24; hncs3_35n12; hncs3_35n24; hncs3_35n9; hncs3_35p17; hncs3_35p22; hncs3_4b4; hncs3_4f17; hncs3_4f5; hncs3_4h6; hncs3_4o19; hncs3_50j4; hncs3_50n06;

hues3_50a23; hues3_6b21; hues3_6c11; hues3_6d16; hues3_72k11; Hues3_72k15;
 hues3_72p16; hues3_7b22; hues3_7d17; hues3_7j3; hues3_7j8; hues3_7p10; hues3_7p9;
 hues3_8e24; Hues3_8g11; Hues3_8g5; hues3_8m10; Hues3_8p7; Hues3_9c22; Hues3_9d20;
 Hues3_9k22; hute1_17k7; hute1_18c12; hute1_18l19; hute1_18u4; hute1_18u1;
 hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2;
 hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15;
 hute1_22d2; hute1_22e12; hute1_22e23; hute1_22e25; hute1_23e13; hute1_23g11;
 hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

2. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16c16; hfr2_16f21; hfr2_16g18; hfr2_16i12; hfr2_16k22; hfr2_16l12; hfr2_22f1; hfr2_22h13; hfr2_22h13; hfr2_22i4; hfr2_22k3; hfr2_22k8; hfr2_23b10; hfr2_23b21; hfr2_23f1; hfr2_23f4; ; hfr2_23n16; hfr2_23o24; hfr2_23o5; hfr2_2a2; hfr2_2b17; hfr2_2b5; hfr2_2c1; hfr2_2c17; hfr2_2c18; hfr2_2d15; hfr2_2d17; hfr2_2d20; hfr2_2g18; hfr2_2h1; hfr2_2h10; hfr2_2i17; hfr2_2k14; hfr2_2k19; hfr2_2c18; hfr2_2f16; hfr2_2g8; hfr2_2j2; hfr2_4lm15; hfr2_62b11; hfr2_62f10; hfr2_62i19; hfr2_62a10; hfr2_62a17; hfr2_64a11; hfr2_64a15; hfr2_64c16; hfr2_64c4; hfr2_64b6; hfr2_64i20; hfr2_64j18; hfr2_64k24; hfr2_64o16; hfr2_6a17; hfr2_6a24; hfr2_6a20; hfr2_6o17; hfr2_71a20; hfr2_72b18; hfr2_72d13; hfr2_72i12; hfr2_72m16; hfr2_72n12; hfr2_78c24; hfr2_78d13; hfr2_78k24; hfr2_78o23; hfr2_7a24; hfr2_7a22; hfr2_7j4; hfr2_82c20; hfr2_10c20; hfr2_82e17; hfr2_10e17; hfr2_82e4; hfr2_10e4; hfr2_82g14; hfr2_10g14; hfr2_82i17; hfr2_10; hfr2_82i24; hfr2_10; hfr2_82m16; hfr2_10; hfr2_82m6; hfr2_10; their complements; and variants thereof.

3. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16f21; hfr2_16k22; hfr2_22f21; hfr2_22h13; hfr2_22i4; hfr2_22k3; hfr2_22k8; hfr2_23f1; ; hfr2_23o24; hfr2_23o5; hfr2_2a2; hfr2_2c1; hfr2_2c18; hfr2_2d20; hfr2_2g18; hfr2_2h1; hfr2_2h10; hfr2_2k19; hfr2_2f16; hfr2_2j2; hfr2_62a10; hfr2_64a11; hfr2_64c16; hfr2_64c4; hfr2_64b6; hfr2_64i20; hfr2_64o16; hfr2_6a17; hfr2_6a20; hfr2_71a20;

hfr2_72d13; hfr2_72m16; hfr2_72n12; hfr2_78d11; hfr2_78n23; hfr2_7a24;
hfr2_7c22; hfr2_7j4; hfr2_82m16; and hfr1_10.

4. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfrd2_1j9; hfrd2_24a15; hfrd2_24b15; hfrd2_24c23; hfrd2_24n20; hfrd2_24p1; hfrd2_3i13; hfrd2_3o17; hfrd2_46a6; hfrd2_46b10; hfrd2_46d13; hfrd2_46j20; hfrd2_46k19; hfrd2_46m4; hfrd2_47a4; hfrd2_46b; hfrd2_4c8; hfrd2_4914; hfrd2_6m11; their complements; and variants thereof.

5. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfrd2_1j9; hfrd2_24c23; hfrd2_46a6; hfrd2_46b10; hfrd2_46d13; hfrd2_46b; hfrd2_4c8; their complements; and variants thereof.

6. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1a11; hmcfl_1c23; hmcfl_1c15; hmcfl_1g13; their complements; and variants thereof.

7. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1c23 hmcfl_1g13; their complements; and variants thereof.

8. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hhes3_1a3; hhes3_14g3; hhes3_14h21; hhes3_14p14; hhes3_14p7; hhes3_15a13; Hhes3_15c24; hhes3_15c6; hhes3_15g14; hhes3_15h1; hhes3_15i5; hhes3_15j18; Hhes3_15j3; hhes3_15k11; hhes3_17f10; hhes3_17f17; hhes3_17n12; hhes3_17n18; Hhes3_18f3; hhes3_18f7; hhes3_19f19; hhes3_19j17; hhes3_1c1; hhes3_1g13; hhes3_1k11; hhes3_20c21; hhes3_20k2; hhes3_20m18; hhes3_21a4; hhes3_21j15; hhes3_21i16; hhes3_21n23; hhes3_22c23; hhes3_22g2; hhes3_22n13; hhes3_23i11; hhes3_23n19; Hhes3_23n19; hhes3_26g22; hhes3_27d1; hhes3_27b4; hhes3_27n14; hhes3_28d14; hhes3_2a11; hhes3_2a17; hhes3_2d15; hhes3_2e12; hhes3_2f14; hhes3_2g7; hhes3_2h1; hhes3_2h15; hhes3_2i19; hhes3_2m18; hhes3_2m20; hhes3_2n9; hhes3_2o4; hhes3_30f4; Hhes3_35b4; hhes3_35h5; hhes3_35e21;

hlex3_35g6; hlex3_35k16; hlex3_35k24; hlex3_35n12; hlex3_35n24; hlex3_35n69;
 hlex3_35p17; hlex3_35p22; hlex3_4b4; hlex3_4f17; hlex3_4f5; hlex3_4h6; hlex3_4o19;
 hlex3_50j4; hlex3_50w06; hlex3_50n23; hlex3_6b21; hlex3_6c11; hlex3_6d16; hlex3_72k11;
 Hlex3_72k15; hlex3_72p16; hlex3_7b22; hlex3_7d17; hlex3_7j3; hlex3_7j8; hlex3_7p10;
 hlex3_7p9; hlex3_8c24; Hlex3_8g11; Hlex3_8g5; hlex3_8m10; Hlex3_8p7; Hlex3_9c22;
 Hlex3_9i20; Hlex3_9k22; their complements; and variants thereof.

9. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hlex3_14g5; hlex3_14p14; hlex3_14p7; hlex3_15a13; hlex3_15g14; hlex3_15h1; hlex3_15j18; hlex3_17f10; Hlex3_18f3; hlex3_19f19; hlex3_19j17; hlex3_20c21; hlex3_21a23; hlex3_22c23; hlex3_22n13; Hlex3_22n19; hlex3_27o14; hlex3_28d14; hlex3_2a11; hlex3_2d15; hlex3_2f14; hlex3_2g7; hlex3_2h15; hlex3_2i19; hlex3_2m20; hlex3_2o9; hlex3_30f4; hlex3_35g6; hlex3_35n24; hlex3_35p17; hlex3_4b4; hlex3_4f17; hlex3_4o19; hlex3_50j4; hlex3_50n23; hlex3_50w06; hlex3_6b21; hlex3_6d16; hlex3_72k11; hlex3_7d17; hlex3_7j8; Hlex3_8g11; Hlex3_8g5; Hlex3_8p7; Hlex3_9c22; Hlex3_9i20; Hlex3_9k22; their complements; and variants thereof.

10. A₁ assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16g18; hfr2_21k14; Hlex3_35b4; hlex3_35p22; hlex3_7j3; hlex3_7p10; hlex3_20n11; their complements; and variants thereof.

11. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16c16; hfr2_21d5; hlex3_15i5; hlex3_18f7; hlex3_1k11; Hlex3_72k15; hlex3_7b22; hlex3_19g22; hlex3_24j6; their complements; and variants thereof.

12. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_2d15; hlex3_35c21; hlex3_2b3; their complements; and variants thereof.

13. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_23i24; hfr2_21i7; hfr2_41m15; hfr2_62f10; hfr2_62i19; hfr2_64j18;

hfxd2_24a20; hfxd2_24p5; hfxd2_4k14; hms3_1g13; hms3_21116; hms3_23111;
hms3_26g22; hms3_4b6; hms3_72p16; hms1_19b17; hms1_20b13; hms1_24e11; their
complements; and variants thereof.

14. An assemblage, comprising at least one nucleic acid molecule having the
sequence of a clone selected from the group consisting of: hfr2_3g8; hfr2_63o17;
hfr2_6b24; hfr2_78b24; hfxd2_24b15; hfxd2_3o17; hfxd2_66p20; hms3_17i17;
hms3_17e18; hms3_27b1; hms3_2a17; hms3_35b5; hms3_35b16; hms3_35n12;
hms3_35e9; hms1_20b19; hms1_20n24; hms1_23c13; their complements; and variants
thereof.

15. An assemblage, comprising at least one nucleic acid molecule having the
sequence of a clone selected from the group consisting of: hfr2_23b10; hfr2_3c18;
hfr2_64a15; hfr2_6o17; hfr2_72b18; hfr2_72i12; hfr2_82i24(hfr1_10);
hms3_14b21; hms3_15j3; hms3_20n18; hms3_22g2; hms3_2m18; hms3_7p9;
hms3_8m10; hms1_18i1; their complements; and variants thereof.

16. An assemblage, comprising at least one nucleic acid molecule having the
sequence of a clone selected from the group consisting of: hfr2_23b21; hfr2_23n16;
hfr2_2c17; hfr2_62b11; hfr2_78c24; hfr2_82a4 (hfr1_10a4); hfr2_82i17
(hfr1_10); hfr2_82m6 (hfr1_10); hfxd2_46m4; hms3_15k11; hms3_1c1; hms3_1a3;
hms3_20k3; hms3_21d4; hms3_23a19; hms3_4f5; hms3_6c11; hms3_8c24; hms1_20g21;
hms1_22d2; hms1_22e12; their complements; and variants thereof.

17. An assemblage, comprising at least one nucleic acid molecule having the
sequence of a clone selected from the group consisting of: hfr2_16i12; hfr2_16i12;
hfr2_22b13; hfr2_2b17; hfr2_2d17; hfr2_64d24; hfr2_82c20 (hfr1_10c20);
hfr2_82e17 (hfr1_10e17); hfr2_82g14 (hfr1_10g14); hfxd2_24a13; hfxd2_3i13;
hfxd2_4m11; hmsf1_1a11; hmsf1_1e13; hms3_15c6; hms3_2a13; hms3_27b4; hms3_2h1;
hms3_33k24; hms1_19f19; and hms1_24c19; their complements; and variants thereof.

18. An assemblage, comprising at least one nucleic acid molecule having the
sequence of a clone selected from the group consisting of: hfxd2_46k19; hfxd2_47a4;

PCT/BR2009/01496

hica3_2c12; hica3_21j15; hica3_17n12; hica1_18i19; hica1_1i2; their complements; and variants thereof.

19. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: huse1_17k; huse1_18c12; huse1_1819; huse1_184; huse1_1811; huse1_19V19; huse1_19g19; huse1_19g22; huse1_19o17; huse1_19j11; huse1_112; huse1_20b19; huse1_20g1; huse1_20a13; huse1_20m11; huse1_20m24; huse1_21a15; huse1_22a2; huse1_22e12; huse1_22a2; huse1_22a2; huse1_23c13; huse1_23g11; huse1_24c19; huse1_24e11; huse1_24j6; huse1_2h3; their complements; and variants thereof.

20. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18d4; hute1_19g19; hute1_19j11; hute1_22n2; hute1_21d15; hute1_22n2; hute1_23g11; their complements; and variants thereof.

[illegible]

hfd2_46010; hfd2_46013; hfd2_46020; hfd2_46019; hfd2_46004; hfd2_4704;
hfd2_4806; hfd2_4608; hfd2_4814; hfd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15;
hmcfl_1g13; hmes3_1a3; hmes3_14g5; hmes3_14h21; hmes3_14p14; hmes3_14p7;
hmes3_15a13; Hmes3_15c24; hmes3_15c6; hmes3_15g14; hmes3_15h1; hmes3_15i5;
hmes3_15j18; Hmes3_15j3; hmes3_15k11; hmes3_17f10; hmes3_17f17; hmes3_17n12;
hmes3_17n18; Hmes3_18f5; hmes3_18f7; hmes3_19f19; hmes3_19j17; hmes3_1e1; hmes3_1g13;
hmes3_1k11; hmes3_20c21; hmes3_20k2; hmes3_20m18; hmes3_21d4; hmes3_21j15;
hmes3_21i16; hmes3_21n23; hmes3_22c23; hmes3_22g2; hmes3_22n13; hmes3_23i11;
hmes3_23n19; Hmes3_23n19; hmes3_26g22; hmes3_27d1; hmes3_27h4; hmes3_27e14;
hmes3_28e14; hmes3_2a11; hmes3_2a17; hmes3_2d15; hmes3_2e12; hmes3_2f14; hmes3_2g7;
hmes3_2h1; hmes3_2h15; hmes3_2i19; hmes3_2m18; hmes3_2m20; hmes3_2n9; hmes3_2o3;
hmes3_30f4; Hmes3_3504; hmes3_35d5; hmes3_35c21; hmes3_35g6; hmes3_35k16;
hmes3_35k24; hmes3_35n12; hmes3_35n24; hmes3_35n9; hmes3_35p17; hmes3_35p22;
hmes3_4b4; hmes3_4f17; hmes3_4f5; hmes3_4h6; hmes3_4o19; hmes3_50f4; hmes3_50n06;
hmes3_50n23; hmes3_6b21; hmes3_6c11; hmes3_6d16; hmes3_72k11; Hmes3_72k15;
hmes3_72p16; hmes3_7b22; hmes3_7d17; hmes3_7j3; hmes3_7j8; hmes3_7p10; hmes3_7p9;
hmes3_8c24; Hmes3_8g11; Hmes3_8g5; hmes3_8m10; Hmes3_8p7; Hmes3_9c22; Hmes3_9c20;
Hmes3_9e22; hme1_17k7; hme1_18c12; hme1_18i19; hme1_18k4; hme1_18l1;
hme1_19f19; hme1_19g19; hme1_19g22; hme1_19h17; hme1_19j11; hme1_1i2;
hme1_20b19; hme1_20g21; hme1_20h13; hme1_20n11; hme1_20m24; hme1_21d15;
hme1_22d2; hme1_22e12; hme1_22n2; hme1_22o2; hme1_23c13; hme1_23g11;
hme1_24c19; hme1_24e11; hme1_24f6; hme1_2h3; their complements; and variants thereof.

22. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:
hfr2_16c16; hfr2_16f21; hfr2_16g18; hfr2_16i12; hfr2_16k22; hfr2_16l12;
hfr2_22f21; hfr2_22h13; hfr2_22h13; hfr2_22i4; hfr2_22k3; hfr2_22k8;
hfr2_23b10; hfr2_23b21; hfr2_23f2; hfr2_23f24; ; hfr2_23e16; hfr2_23e24;
hfr2_23e5; hfr2_2a2; hfr2_2b17; hfr2_2b5; hfr2_2c1; hfr2_2c17; hfr2_2c18;
hfr2_2d15; hfr2_2d17; hfr2_2d20; hfr2_2g18; hfr2_2h1; hfr2_2h10; hfr2_2i17;
hfr2_2k14; hfr2_2k19; hfr2_2k18; hfr2_2f16; hfr2_2g8; hfr2_3j2; hfr2_41m15;
hfr2_62b11; hfr2_62f10; hfr2_62i19; hfr2_62n10; hfr2_62o17; hfr2_64a11;

hfr2_64a15; hfr2_64c16; hfr2_64c4; hfr2_64b6; hfr2_64d20; hfr2_64j18;
 hfr2_64k24; hfr2_64o16; hfr2_6a17; hfr2_6b24; hfr2_6d20; hfr2_6o17;
 hfr2_71a20; hfr2_72b18; hfr2_72d13; hfr2_72i12; hfr2_72m16; hfr2_72n12;
 hfr2_78c24; hfr2_78d13; hfr2_78k24; hfr2_78a23; hfr2_7a24; hfr2_7c22;
 hfr2_7j4; hfr2_83c20; hfr2_10c20; hfr2_82e17; hfr2_10e17; hfr2_83a4;
 hfr2_10e4; hfr2_82g14; hfr2_10g14; hfr2_82i17; hfr2_10i; hfr2_82j24; hfr2_10;
 hfr2_82m16; hfr2_10; hfr2_82n6; hfr2_10; complements of the nucleic acid
 sequences; and variants thereof.

23. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:
 hfr2_16d21; hfr2_16k22; hfr2_22f21; hfr2_22n13; hfr2_22i4; hfr2_22k3; hfr2_22k8;
 hfr2_23f2; hfr2_23a24; hfr2_23a5; hfr2_2a2; hfr2_2e1; hfr2_2e18; hfr2_2d20;
 hfr2_2g18; hfr2_2b1; hfr2_2b10; hfr2_2k19; hfr2_2f16; hfr2_2i2; hfr2_62n10;
 hfr2_64a11; hfr2_64c16; hfr2_64a4; hfr2_64b6; hfr2_64d20; hfr2_64k24;
 hfr2_64a16; hfr2_6a17; hfr2_6d20; hfr2_71a20; hfr2_72d13; hfr2_72m16;
 hfr2_72n12; hfr2_78d13; hfr2_78a23; hfr2_7a24; hfr2_7c22; hfr2_7j4; hfr2_82m16;
 hfr2_10; complements of the nucleic acid sequences; and variants thereof.

24. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:
 hfrd2_1j9; hfrd2_24a15; hfrd2_24b15; hfrd2_24e23; hfrd2_24n20; hfrd2_24p5;
 hfrd2_3i13; hfrd2_3o17; hfrd2_46a6; hfrd2_46b10; hfrd2_46d13; hfrd2_46j30;
 hfrd2_46k19; hfrd2_46m4; hfrd2_47a4; hfrd2_4b6; hfrd2_4c8; hfrd2_4k14;
 hfrd2_4m11; complements of the nucleic acid sequences; and variants thereof.

25. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of: hfrd2_1j9;
 hfrd2_24e23; hfrd2_46a6; hfrd2_46b10; hfrd2_46d13; hfrd2_4b6; hfrd2_4c8;
 complements of the nucleic acid sequences; and variants thereof.

26. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:

hmcfl_1a11; hmcfl_1c23; hmcfl_1c15; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

27. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmcfl_1c23; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

28. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hhes3_1a3; hhes3_14g5; hhes3_14b21; hhes3_14p14; hhes3_14p7; hhes3_15a13; Hhes3_15c24; hhes3_15c6; hhes3_15g14; hhes3_15h1; hhes3_15i5; hhes3_15j18; Hhes3_15j3; hhes3_15k11; hhes3_17f10; hhes3_17i17; hhes3_17n12; hhes3_17n18; Hhes3_18f; hhes3_18f7; hhes3_19f19; hhes3_19j17; hhes3_1c1; hhes3_1g13; hhes3_1k11; hhes3_20c21; hhes3_20c2; hhes3_20n18; hhes3_21d4; hhes3_21j15; hhes3_21i16; hhes3_21n23; hhes3_22c23; hhes3_22g2; hhes3_22n13; hhes3_23i11; hhes3_23n19; Hhes3_23n19; hhes3_24g22; hhes3_27d1; hhes3_27d4; hhes3_27o14; hhes3_28d14; hhes3_2a11; hhes3_2a17; hhes3_2d15; hhes3_2e12; hhes3_2f14; hhes3_2g7; hhes3_2h1; hhes3_2h15; hhes3_2i19; hhes3_2m18; hhes3_2m20; hhes3_2n9; hhes3_2o3; hhes3_30f4; Hhes3_35b4; hhes3_35b5; hhes3_35c21; hhes3_35g6; hhes3_35k16; hhes3_35k24; hhes3_35n12; hhes3_35n24; hhes3_35n9; hhes3_35p17; hhes3_35p22; hhes3_4b4; hhes3_4f17; hhes3_4f5; hhes3_4b6; hhes3_4o19; hhes3_50f4; hhes3_50n6; hhes3_50n23; hhes3_6d21; hhes3_6c11; hhes3_6d16; hhes3_72k11; Hhes3_72k15; hhes3_72p16; hhes3_7b22; hhes3_7d17; hhes3_7j3; hhes3_7j8; hhes3_7p10; hhes3_7p9; hhes3_8c24; Hhes3_8g11; Hhes3_8g5; hhes3_8n10; Hhes3_8p7; hhes3_9e22; Hhes3_9i20; Hhes3_9c22; complements of the nucleic acid sequences; and variants thereof.

29. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hhes3_14g5; hhes3_14p14; hhes3_14p7; hhes3_15a13; hhes3_15g14; hhes3_15h1; hhes3_15j18; hhes3_17f10; hhes3_17n18; Hhes3_18f; hhes3_19f19; hhes3_19j17; hhes3_20c21; hhes3_21n23; hhes3_22c23; hhes3_22n13; Hhes3_23n19; hhes3_27o14; hhes3_28d14; hhes3_2a11; hhes3_2d15; hhes3_2f14; hhes3_2g7; hhes3_2h15; hhes3_2i19; hhes3_2m20; hhes3_2n9; hhes3_30f4; hhes3_35g6; hhes3_35n24; hhes3_35p17; hhes3_4b4; hhes3_4f17;

hncs3_4o19; hncs3_5q4; hncs3_5o23; hncs3_5o66; hncs3_6o21; hncs3_6d16; hncs3_72k11;
hncs3_7d17; hncs3_7j8; hncs3_8g11; hncs3_8g5; hncs3_8p7; hncs3_9e22; hncs3_9o20;
hncs3_9e22; complements of the nucleic acid sequences; and variants thereof.

30. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2_16g18; hfor2_2h14; hncs3_35p4; hncs3_35p22; hncs3_7j3; hncs3_7p10; hncs3_7o22; complements of the nucleic acid sequences; and variants thereof.

31. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2_16c16; hfor2_2b5; hncs3_15i5; hncs3_18i7; hncs3_1k11; hncs3_72k15; hncs3_7b22; hncs3_19g22; hncs3_34j6; complements of the nucleic acid sequences; and variants thereof.

32. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2_2d15; hncs3_35c21; hncs3_2h3; complements of the nucleic acid sequences; and variants thereof.

33. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2_23i24; hfor2_2i17; hfor2_41m15; hfor2_62f10; hfor2_62i19; hfor2_64j18; hfor2_24c20; hfor2_24p5; hfor2_4k14; hncs3_1g13; hncs3_21i16; hncs3_23i11; hncs3_76p22; hncs3_4b6; hncs3_72p16; hncs3_19h17; hncs3_20h13; hncs3_24c11; complements of the nucleic acid sequences; and variants thereof.

34. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2_3g8; hfor2_62o17; hfor2_6624; hfor2_78k24; hfor2_24b15; hfor2_3o17; hfor2_46p20; hncs3_17i17; hncs3_17a18; hncs3_27d1; hncs3_2a17; hncs3_35b5; hncs3_35k16; hncs3_35m12; hncs3_35p9; hncs3_20b19; hncs3_20m24; hncs3_23e13; complements of the nucleic acid sequences; and variants thereof.

35. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfr2_23b10; hfr2_3c18; hfr2_64a15; hfr2_6c17; hfr2_72b18; hfr2_72b12;
hfr2_82d24(hfr1_10); hms3_14b11; hms3_15j3; hms3_20m18; hms3_22g2; hms3_2m18;
hms3_7p9; hms3_8m10; hms1_181; complements of the nucleic acid sequences; and
variants thereof.

36. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfr2_23b21; hfr2_23n16; hfr2_2c17; hfr2_62b11; hfr2_78c24; hfr2_82c4
(hfr1_10c4); hfr2_82i17 (hfr1_10); hfr2_82m6 (hfr1_10); hfr2_46m4; hms3_15k11;
hms3_1c1; hms3_1a3; hms3_20k2; hms3_21d4; hms3_23n19; hms3_4f5; hms3_6c11;
hms3_8c24; hms1_20g11; hms1_22d2; hms1_22e12; complements of the nucleic acid
sequences; and variants thereof.

37. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfr2_16i12; hfr2_16i12; hfr2_22b13; hfr2_2b17; hfr2_2d17; hfr2_64d24;
hfr2_82c20 (hfr1_10c20); hfr2_82e17 (hfr1_10e17); hfr2_82g14 (hfr1_10g14);
hfr2_24a15; hfr2_3i13; hfr2_4m11; hmsf1_1a11; hmsf1_1c15; hms3_15c8;
hms3_2a3; hms3_27k4; hms3_2b1; hms3_33c24; hms1_19f19; and hms1_24c19;
complements of the nucleic acid sequences; and variants thereof.

38. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfr2_46k19; hfr2_47a4; hms3_2e12; hms3_21j15; hms3_17a12; hms1_18i19;
hms1_1i2; complements of the nucleic acid sequences; and variants thereof.

39. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hms1_17k7; hms1_18c12; hms1_18i19; hms1_18d4; hms1_18i1; hms1_19f19;
hms1_19g19; hms1_19g22; hms1_19h17; hms1_19j11; hms1_1i2; hms1_20b19;
hms1_20g21; hms1_20h13; hms1_20m11; hms1_20m34; hms1_21d15; hms1_22d2;
hms1_22e12; hms1_22n2; hms1_22o2; hms1_23e13; hms1_23g11; hms1_24c19;
hms1_24e11; hms1_24f6; hms1_25b3; complements of the nucleic acid sequences; and
variants thereof.

40. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: huse1_17k7; huse1_18c12; huse1_18i4; huse1_19g19; huse1_19j11; huse1_22n2; huse1_21d15; huse1_22o2; huse1_23g11; complements of the nucleic acid sequences; and variants thereof.

41. A nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2_16c16; hfr2_16f21; hfr2_16g18; hfr2_16i2; hfr2_16k22; hfr2_16l12; hfr2_22f21; hfr2_22h13; hfr2_22h13; hfr2_22i4; hfr2_22k3; hfr2_22k8; hfr2_23b10; hfr2_23b21; hfr2_23f2; hfr2_23i4; hfr2_23n16; hfr2_23o24; hfr2_23o5; hfr2_2a2; hfr2_2b17; hfr2_2b5; hfr2_2c1; hfr2_2c17; hfr2_2c18; hfr2_2d13; hfr2_2d17; hfr2_2d20; hfr2_2g18; hfr2_2h1; hfr2_2h10; hfr2_2i17; hfr2_2k14; hfr2_2k19; hfr2_2k16; hfr2_2c18; hfr2_2f16; hfr2_2g8; hfr2_2i2; hfr2_41m15; hfr2_62h11; hfr2_62f10; hfr2_62i19; hfr2_62n10; hfr2_62o17; hfr2_64a11; hfr2_64a15; hfr2_64c16; hfr2_64c4; hfr2_64b6; hfr2_64i20; hfr2_64j18; hfr2_64k24; hfr2_64o16; hfr2_6a17; hfr2_6b24; hfr2_6d20; hfr2_6o17; hfr2_71c20; hfr2_72b18; hfr2_72d13; hfr2_72i13; hfr2_72m16; hfr2_72n12; hfr2_78c24; hfr2_78d13; hfr2_78k24; hfr2_78n23; hfr2_7a24; hfr2_7e22; hfr2_7f4; hfr2_82c20; hfr2_10c20; hfr2_82e17; hfr2_10e17; hfr2_82e4; hfr2_10e4; hfr2_82g14; hfr2_10g14; hfr2_82i17; hfr2_10; hfr2_82i24; hfr2_10; hfr2_82m16; hfr2_10; hfr2_82m6; hfr2_10; hfr2_109; hfr2_24a15; hfr2_24b15; hfr2_24c23; hfr2_24d20; hfr2_24p5; hfr2_2i13; hfr2_2j17; hfr2_46a6; hfr2_46b10; hfr2_46d13; hfr2_46i20; hfr2_46k19; hfr2_46m4; hfr2_47a4; hfr2_47b6; hfr2_4c8; hfr2_4h14; hfr2_4m11; huse1_1a11; huse1_1c23; huse1_1e15; huse1_1g13; huse1_1h3; huse1_14p5; huse1_14b21; huse1_14p14; huse1_14p7; huse1_15a13; huse1_15c24; huse1_15c5; huse1_15g14; huse1_15h1; huse1_15i5; huse1_15j18; huse1_15j3; huse1_15k11; huse1_17f10; huse1_17i17; huse1_17n12; huse1_17n18; huse1_18f5; huse1_18i7; huse1_19f19; huse1_19j17; huse1_1c1; huse1_1g13; huse1_1k11; huse1_20c21; huse1_20k2; huse1_20m18; huse1_21d4; huse1_21j15; huse1_21i16; huse1_21n23; huse1_22c23; huse1_22g2; huse1_22n13; huse1_23i11; huse1_23n19; huse1_23n23; huse1_23g22; huse1_27d1; huse1_27k4; huse1_27o14; huse1_28d14; huse1_2a11; huse1_2a17; huse1_2d15; huse1_2e12; huse1_2f14; huse1_2g7; huse1_2h1; huse1_2h15; huse1_2i19; huse1_2m18;

hues3_2m20; hues3_2m9; hues3_2m3; hues3_304; Hues3_35b4; hues3_35b5; hues3_35b21;
 hues3_35g6; hues3_35h16; hues3_35h24; hues3_35h12; hues3_35h24; hues3_35h9;
 hues3_35p17; hues3_35p22; hues3_4b4; hues3_4f17; hues3_4f5; hues3_4b6; hues3_4c19;
 hues3_504; hues3_50a06; hues3_50a23; hues3_6a21; hues3_6c11; hues3_6d16; hues3_72h11;
 Hues3_72h15; hues3_72p16; hues3_7b22; hues3_7d17; hues3_7j3; hues3_7j8; hues3_7p10;
 hues3_7p9; hues3_8c24; Hues3_8g11; Hues3_8g5; hues3_8m10; Hues3_8p7; Hues3_9c22;
 Hues3_9c20; Hues3_9c22; hues1_17h7; hues1_18c12; hues1_18i19; hues1_18a4; hues1_18i1;
 hues1_19f19; hues1_19g19; hues1_19g22; hues1_19h17; hues1_19j11; hues1_1i2;
 hues1_20b19; hues1_20g21; hues1_20h13; hues1_20m11; hues1_20m24; hues1_21d15;
 hues1_22d2; hues1_22e12; hues1_22a2; hues1_22o2; hues1_23e13; hues1_23g11;
 hues1_24c19; hues1_24c11; hues1_24f6; hues1_2h3; their complements; and variants thereof.

42. A polypeptide encoded by the nucleic acid molecule according to claim 41.
43. An antibody or fragment thereof that is capable of binding to a specific portion of the peptide according to claim 42.
44. A pharmaceutical composition, comprising (a) an effective amount of a pharmaceutical agent, wherein said pharmaceutical agent is selected from the group consisting of the polypeptide according to claim 42, variants or functional derivatives thereof, and antibodies thereon; and (2) a physiologically acceptable carrier or excipient.
45. An expression vector comprising the nucleic acid molecule of claim 41 or a fragment thereof, and optionally a promoter operably linked to said nucleic acid molecule or said fragment.
46. A method for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by the nucleic acid molecule according to claim 41.